UNDERSTANDING THE DYNAMIC NATURE OF TEACHER/CLASSROOM EFFECTS ON EDUCATIONAL OUTCOMES: A CROSS-CULTURAL INVESTIGATION

Elaine Karyn White

Department of Psychology
Goldsmiths University of London

Thesis submitted for the degree of Doctor of Philosophy
March 2017

Declaration

I declare that the work presented in this thesis is my own. All experiments and work detailed in the text of this thesis is novel and has not been previously submitted as part of the requirements of a higher degree.

Signed ______________________________________ Date ___________

Abstract

The idea that teachers differ substantially in their ability to motivate and educate students has pervaded educational research for decades. While the education system, and teachers in particular, provide an enormously important service, many people hold teachers almost entirely responsible for differences between classes and for individual students' performance. The belief that the 'teacher effect' is such that students would perform better or worse given a specific teacher remains unfounded, as true experimental design is difficult to apply. The present thesis, employing pseudo-experimental methods, investigated potential teacher/classroom effects on several educational outcomes. The five empirical chapters in this thesis explored whether students' motivation, academic performance, and perception of learning environment were affected by their teachers and/or classmates, as reflected in average differences between classes. Investigations were conducted longitudinally and cross-culturally, in three different education systems using data from four samples. Two samples were secondary school students aged 10 to 12 years, in their first year of secondary education, from the UK and Russia, and two samples were large representative developmental twin studies, the Twins Early Development Study (TEDS) from the UK, and the Quebec Newborn Twin Study (QNTS) from Quebec, Canada. Average differences were observed across classrooms and teacher groups, effect sizes ranging from 2% to 25%. The results suggested a weak influence of current subject teacher that was difficult to disentangle from several confounding factors, such as peer influences, selection processes, individual differences in ability and perceptions, teacher characteristics and evocative processes. The findings suggest that student

outcomes, rather than being predominantly influenced by teacher effects, are under multiple influences. Overall, the results call for caution in considering ‘added value’ or ‘teacher effect’ measures as valid criteria for current education policies that affect teacher promotion and employment prospects.

Statement of Authorship

The data collected from UK secondary school students were collected across five assessments during one academic year. In the UK, I was personally responsible for recruiting participants and coordinating data collections. Several students from Goldsmiths, University of London assisted with each collection wave as testing took place within students' maths classes, and therefore data were collected from four classes of students, simultaneously.

Data collected from Russian secondary school students were collected on my behalf by students and researchers from the Laboratory for Cognitive Investigations and Behavioural Genetics, Tomsk State University, Tomsk in conjunction with teachers of the schools.

The data from the Twins' Early Development Study (TEDS) were collected as part of a large collaborative project funded by the UK Medical Research Council. Collaborators at University Laval, Quebec, Canada, collected the data from the Quebec Newborn Twin Study (QNTS).

I personally conducted all data analyses reported in the present thesis, apart from data reported in Chapter 7 from the QNTS that were analyzed on my behalf by Gabrielle Garon-Carrier from University Laval, Quebec, Canada, who is also joint co-author of this study. A slightly modified version of this Chapter has been submitted to the journal *Developmental Psychology.*

The work presented in this thesis is original and my own.

Acknowledgements

First and foremost, I would like to extend enormous gratitude to my first supervisor, Professor Yulia Kovas who has been a complete inspiration and provided tremendous support and guidance throughout my PhD, and since my undergraduate years. Enormous gratitude also extends to my second supervisor, Dr. Alice Jones who has also been an inspiration and provided unwavering support and guidance since the beginning of my academic journey. I would like to thank you both for sharing your knowledge and for your belief in me.

My gratitude also extends to Tatiana Kolienko, Dr. Efrosini Setakis and Mahnaz Tavakoly, without whom the project would not have been possible. I am also extremely grateful to Ekaterina Kolienko for her exceptional work on the Russian data input. I am most grateful to the participating schools, teachers and students who allowed us to disrupt their maths and geography lessons to collect data. I would also like to thank the TEDS and QNTS twins and their families for their participation.

I am incredibly grateful to my amazing friends and colleagues who not only assisted in the planning process, data collection, and data input, at InLab, but have also been a fantastic support (Kaili Rimfeld, Ekaterina Cooper, Maria Grazia Tosto, Maja Rodic, Margherita Malanchini, Robert Chapman, Emma Martin, Felicia Francis-Stephenson, Nicholas Kendall, Saba Hassan, Maria Rudenko, Britta Schunemann, Kathy Filer and Liam Broom) and at the Laboratory for Cognitive Investigations and Behavioural Genetics, Tomsk State University, Tomsk, (Olga Bogdanova and Dina Zueva). I am also extremely grateful to Gabrielle Garon-Carrier, thank you for your support and guidance.

I would also like to extend my gratitude to Professor Sergey Malykh, and Professor Michel Boivin at the Laboratory for Investigations into Biopsychosocial Factors in Child Development, Tomsk State University for their support, without which my PhD journey would not have been possible.

Finally, I would like to say special thanks to my husband, Rob, my mum, Sylvia, my brother Martin, and my sister-in-law, Joanne, thank you for your love and enormous support always, this thesis is dedicated to you all.

Table of Contents

List of Figures

Chapter 3

Chapter 5

Chapter 6

Chapter 7

List of Tables

Chapter 5, Part 5.1

Chapter 5, Part 5.2

Chapter 6

Chapter 7

Author’s Publications

White, E., Davydova, Y. A., Sharafieva, K. R., Malykh, S. B. & Kovas, Y. V. (2012). The individual differences in spatial abilities: An investigation into age, sex, handedness and sibling effects. *Theoretical and Experimental Psychology*

Voronina, I. D., Bohan, T.G., Terehina, O.V., **White, E.,** Malykh, S.B., Kovas. Y.V. (2016). Demographics, lifestyle and health in families with natural and induced pregnancy in Russia and UK. *Theoretical and Experimental Psychology*

White, E. K., Garon-Carrier' G., Tosto, M. G., Malykh, S. B., Li, X., Tick, B., Riglin, L., Byrne, B. Dionne, G., Brendgen, M., Vitaro, F., Tremblay, R. E., Boivin, M., & Yulia Kovas, Y. (Sumitted). Twin classroom dilemma: to study together or separately? *Developmental Psychology*

Chapter 1

Introduction

The education system, and teachers in particular, provide an enormous service but there seems to be much negativity aimed at them especially concerning their students' school achievement (Kovas, Malykh, Gaysina, 2016). It appears that schools, and teachers in particular, receive little praise when students do well, but receive much criticism when students appear to underachieve (Christodoulu, 2014; Olson, Keenan, Byrne, & Samuelsson, 2014). This is especially so for maths achievement and becomes evident when the Programme for International Student Assessment (PISA) results are announced and the news media highlight how far the UK is lagging behind in maths globally (e.g. Coughlan, 2016). PISA publishes reports every four years showing mean differences between the participating countries in students' maths, reading and science performance. The reports consistently show East Asian countries like China, Singapore and Korea as the top performers (OECD, 2010; 2009; 2012, 2016). In contrast, the UK, US and Russia appear to be around average. This has become a huge topic for educational policy makers as they try to increase maths performance in their populations. Despite mean differences in maths performance, the differences between countries are small compared to variation within countries, with around 90% of this variation overlapping across countries (OECD, 2009). Most of the variation is seen within countries, schools and classrooms (Kovas, et al., 2007). Often these individual differences are overlooked in educational research

One area of the educational field where individual differences are not often accounted for is classroom or teacher effects research. Classroom and/or teacher effects are translated as measures of teacher/classroom effectiveness in relation to variation in student outcomes (e.g. Nye, Konstantopoulos & Hedges, 2004). The emphasis is usually on the teacher with the assumption being that average achievement of students in a given class would be higher or lower if they had a different teacher (e.g. Hanushek, 2011). These are average effects, so they may, or may not apply to each individual student, although the common assumption is that they do (Loeb, 2013). Variation in student achievement is influenced by numerous factors apart from teachers and classrooms, such as motivation (Spinath, Spinath, Harlaar, & Plomin, 2006), and ability (Deary, Strand, Smith, & Fernandes, 2007). Therefore when considering the classroom, individual student as well as teacher characteristics need to be taken into account. The learning environment is multifaceted, so a more comprehensive and detailed approach to its research is necessary.

To date, there has been a large body of research investigating teacher\classroom effects but the results have been limited due to the broad approach taken by most studies. This review evaluates key research into teacher/classroom effects in relation to student achievement and motivation. The review considers different approaches, including large-scale survey studies, random allocation and behavioural genetics research. It also assesses research that studied other factors often reported to influence student achievement. These include investigations into class size, classroom composition/streaming, and teacher characteristics. The research shows some average effects of the teacher or class but the sizes of these effects are modest. On average, the effect of being in a particular class has been shown to be around 8% (e.g. Byrne et al., 2010).

Studies often base their findings on student achievement gains, which are calculated from students' test scores. In the US, where many studies are conducted, the Stanford Achievement Test (SAT) scores are frequently used. The SAT is a standardised achievement test taken yearly by students across US school districts. The annual completion of the test allows gains in achievement to be calculated across years of education. However, if as suggested, the SAT is not necessarily curriculum specific then gains on this test cannot be inferred as genuine school achievement (Konstantopoulos (2008). This should be cause for concern, considering these achievement gains are frequently used as an index of teacher quality (e.g. Chetty, Friedman & Rockoff, 2011; Nye et al., 2004).

Several studies take an econometric approach whereby they use data in large-scale samples of around 2.5 million. Local and national authorities have usually gathered the data on their populations for administration purposes, so while it is large numerically, it lacks detailed information. Consequently, these large-scale survey studies are useful to explore simple relationships but are unable to account for other factors within the classroom that likely contribute towards student achievement (Samuelson, 2004). For example, factors such as educational processes, ability selection and individual student characteristics need to be considered when examining teacher/classroom effects.

Large Scale Studies Of Teacher Effects

Random Allocation

One such study that investigated teacher/classroom effects using data from whole districts in the US was Project STAR (Student-Teacher Achievement Ratio). It was established in 1985 as a collaborative endeavour between Tennessee State Legislature, Tennessee State Department of Education and a consortium of Tennessee universities (Nye et al., 2004). In order to empirically investigate teacher and class size effects in relation to student achievement, a design with a strong experimental control was applied, whereby 4,000 students and their teachers were randomly assigned to their classes. The sample was followed for four years, commencing when students were aged 5 years through to age 9. Although modest effects of within-school between-teacher variance were found across grades for reading (ranging from 6.6 to 7.4%) and maths (ranging from12.3 to 13.5%), greater effects were shown for low compared with high socio-economic status (SES) schools in 3rd grade reading (14% low SES and 3.8% high SES). Teacher experience was also shown to predict achievement gains in 2nd grade reading and 3rd grade maths of 0.15 standard deviations (SD) and 0.19 SD respectively. Despite the scale and breadth of the study, and the ability to control for class size effects, there are several limitations. For instance, the study has been criticised for not being 'blind' in that students, teachers, parents and educational personnel were all aware of the programme (Hanushek, 1999). This may have had some influence on results in terms of motivation and resources. Although a broad range of schools took part, it is unknown whether or not schools implemented ability selection or streaming. There were also a small number of students who moved between classes and may have confounded class-size effects. Although an effect of teacher experience was found, it was unclear how and when this was measured. Another detracting factor is that pre-randomisation test scores were not collected so achievement gains were obtained by comparing against previous year test scores within the study. Further, the effect shown for low SES schools may not be due to teacher effects, as these results are consistent with research showing greater effects of high quality pre-school child-care on low income compared with high income children (e.g. Caughy, DiPietro, & Strobino,1994; Dearing, McCartney, & Taylor, 2009; Geoffroy, Côté, Borge, Larouche, Séguin, & Rutter, 2007). If these effects were due to teacher effects per se, they would have likely been shown for all SES groups. Instead, they likely show the impact of learning environment on lower cognitive scores at baseline for the low SES schools as seen in low SES children who were entered into pre-school childcare earlier than average. The children benefited such that they outperformed their higher SES peers on tests at age 5 and 6 when otherwise they would have been subject to reduced school readiness (Caughy et al., 1994).

Teacher Quality

Another way to assess teacher effects is to use a 'teacher value added' (VA) measure. Also known as the impact teachers have on their students' test scores, VA is estimated from student achievement gains (Loeb, 2013). VA was examined in another large-scale survey study across an urban US district with a sample of over 970,000 students (Chetty, Friedman, & Rockoff, 2011). Student achievement gains were measured in maths and English test scores across six grades 3 to 8 (ages 8 to 14 years) to explore long-term effects of VA on SES in adulthood. The study controlled for selection bias, a known confounder of teacher VA, by accounting for teacher assignment to schools/classes and parental characteristics such as SES. From these data the study concluded there was no evidence of selection bias for VA. Students with higher VA teachers showed better long-term outcomes. They were less likely to become teenaged parents and more likely to continue their education beyond compulsory schooling, attend higher standard colleges and demonstrate higher levels of neighbourhood SES and earnings. A 1 SD improvement in teacher VA over the course of one academic year led to a 0.1SD increase in test scores averaged across maths and English (0.118 SD for maths and 0.081 for English). A similar relationship was shown for earnings with 1SD improvement in VA translating to an average earnings increase of 0.9% at age 28. A 1SD test score increase also associated with 1% teenage birth decrease and a rise in neighbourhood SES by 1.44%. However, these effects appear to be modest, especially in relation to college attendance at age 20 of 0.49% in relation to 1 SD VA. For earnings, when applied to the average US salary of 41,673.83 US dollars in 2010 (last year of tax data used in the study), a 1% increase would only mean an additional 417 dollars over the course of a year. It is also unclear whether the reported 0.9% increase in average earnings is yearly or monthly. It is possible that the observed effects are the result of false positive significance associated with such a large sample size (Bentler & Bonnett, 1980) rather than any real effect of teacher value added.

The economic value of student outcomes was also assessed in a review of teacher quality (Hanushek, 2011). Standard deviations (SD) of student achievement were used to measure teacher effectiveness along with future earnings of students. Several longitudinal studies were reviewed that considered within-school variance in teacher effectiveness estimated from student achievement gains. The review found average effects of 0.17 SD for maths and 0.13 SD for reading. In secondary analyses of student cognitive skills in relation to young adult earnings, the review suggests that 1 SD increase in maths achievement at the end of high school leads to 10-15% increase in annual salary. It concludes that teacher effectiveness 1 SD above the mean, generated student earning gains of above 400,000 US dollars for a class of 20 students and that US maths and science rankings could potentially be improved by replacing lower quality with average quality teachers. The implication being that a teacher has the same impact across all students within a class. Differences between teachers were also reported whereby some teachers' classes repeatedly had larger gains equivalent to 1.5 years' achievement gains while others with similar students had recurrent gains of only 0.5 year. The review also suggested that the impact of teacher on student achievement is far greater than any other attribute of the school. While teachers are undeniably important, no consideration is given to other within-classroom factors such as peer effects that may also influence student achievement (Burke & Sass, 2013). It is also unclear whether schools and classes in the studies reviewed were subject to ability selection, as this would also lead to differences between classes and teachers in achievement gains. This study is one example of teacher effect research that uses classroom performance as an index of teacher quality. Considering the numerous other factors that contribute towards student achievement, it is concerning that teacher employment prospects are based on classroom performance.

Other research suggests that while student achievement gains should be an ideal measure of teacher effect, there is little variance among students in real academic growth, and so they provide an inaccurate assessment of differences between individual students' rates of change (Rowan, Correnti & Miller, 2002). In an evaluation of large-scale survey studies, it was suggested that methodologies were in need of improvement (Rowan et al., 2002). Using a large elementary school data set from Prospects: The Congressionally Mandated Study of Educational Opportunity (Rowan et al., 2002), the study examined differences in the magnitude of effects found in previous research that investigated teacher value added. It also proposed that as measurement error is accounted for at the student level this might lead to a reduction in teacher effect coefficients and therefore underestimate teacher/class effects on achievement. As part of the evaluation of previous methodologies, the study first conducted cross-sectional analyses using adjusted student achievement gains whereby achievement in a year is used as a criterion variable in regression analyses while controlling for other factors such as previous attainment, home and school SES. Similarly to other such studies, small effects were found for between-classroom variance for maths and reading (6-13% and 3-10% respectively). Using cross-classified random effects models, a new approach at the time of publishing, much larger effects were found for between student variance of 27-28% - reading and 13-19% - maths. Further, when controlling for prior achievement, home and school SES, classrooms explained ~60-61% of variance in growth for reading attainment and ~52-72% of the variance in growth in maths attainment, effect sizes (d) ranged from .77 to .78 for reading and .72 to .85 for maths. The study also found consistency among different subjects and grades for these estimates. More variation was found depending on background factors; for example, attainment growth was not equal across SES within the same school. While the study reported variation in student growth (estimated as d=0.07), it is unclear what actual growth occurred to evaluate the observed effects. The study does, however, take account of other factors that contribute towards student achievement. Additionally, the authors caution that while large classroom effects have been found, what constitutes effective teaching/class environments is largely unknown. They suggest using intervention studies that manipulate teaching practices, such as class size, class composition, or streaming practices.

Potential Sources Of Influence On Student Achievement

Class Size

Random allocation

Several intervention studies have manipulated class size as a potential source of classroom effects. The optimum number of students within one classroom has been under considerable debate in relation to student achievement. The findings in the literature are mixed regarding whether reducing class size increases student achievement. Overall, any effects revealed are small and reducing class size does not necessarily help the most disadvantaged groups. One study in particular, Project STAR (Student-Teacher Achievement Ratio) began in 1985 (Nye, Hedges & Konstantopoulos, 2000; Konstantopolous, 2008) using a design where 6,840 students and their teachers were randomly assigned to classes of different sizes. The sample was followed for four years, from kindergarten to 3rd grade (aged 5 years at commencement). The project, a large-scale survey study, invited all schools in Tennessee to take part. Many smaller schools were unable to participate, as with fewer classes at each grade it was not possible to use the randomized design. A total of 79 schools in 42 school districts participated in the study that consisted of extra testing for students; researchers visiting to verify class size; interviews; and data collection. Additional teachers were recruited and paid for by the project to enable the extra classes. Three types of classes were devised, small classes of 13-17 students with one teacher; large classes with one teacher and 22-26 students; and large classes (22-26) with one teacher plus a full-time classroom assistant. Nye and colleagues reported positive effects of small classes on maths and reading of between 0.15 to 0.30 SD units, with stronger influence on the early grades. Further analyses found this effect to be greater for higher achieving students compared with other students. For example, in kindergarten mathematics, for students in the 90th quantile the effect was twice that of students in the 50th quantile, and four times students in the 10th quantile. These effects were significantly different at the $p \leq .05$ level (Konstantopoulos, 2008). Despite, the scope of the project, it has been subject to criticism for some methodological limitations. For example, there may be a degree of selection bias in the sample. The commitment to take part for four years and provide the additional numbers of classrooms would likely restrict the sample to those better resourced. This may mean that participating schools were more likely to be those already doing better. Additionally, the schools had to accommodate the necessary extra classrooms (e.g. Goldstein & Blatchford, 1998). Another limitation is that some students switched between classes during the study. It also appears that on average, the sizes of the smaller classes were adhered to; but for the larger classes, the average size was 22 and some class sizes overlapped with the smaller classes. Lack of clear distinction between the class sizes would likely undermine any conclusions of the study. The project was also subject to considerable attrition across the term of the study, although Nye and colleagues suggest that this made little or no difference to their results when comparing pre and post-dropout treatment effects for these students. They maintain that in terms of policy making, small classes appear to be more beneficial for achievement than larger ones. However, this effect did not extend to underachievers. Further, as with other large-scale survey studies, the specific reasons of why smaller classes might be advantageous is unknown.

Another study that used a random allocation approach examined class size in relation to literacy in the French education system (Ecalle, Magnan & Gibert, 2006). Students and teachers were randomly assigned to either experimental small classes of 10 to 12 students or regular classes (20-25 students). Schools were randomly selected from those with special educational needs, with lower test scores for at least half of their students. The students (*N*=1,192) were aged 6 to 7 years, including those at normal reading age and those who were repeating the year to catch up. The study took initial baseline measures of literacy at the beginning of the school year; these were used as covariates along with other factors including age, early schooling, first language and SES. Two further assessments were conducted in the spring and summer. The results showed that class size had a modest effect on literacy performance of 1% with better performance in the smaller classes. However, it was students with French as their first language, i.e. the most advantaged, in the small classes that made the most progress. Those with French as their second language had equivalent performance to their counterparts in the control group. It appeared that the intervention made little or no difference to the more disadvantaged groups that it was hoped would benefit the most. Furthermore, 28 to 29% of the variance in literacy performance was accounted for by students' initial literacy skills, suggesting that class size had negligible influence above student ability. These results also suggest that students with the most severe special educational needs require more assistance than just a smaller class size. Additionally, only limited conclusions can be made from studies that control just one feature of school, without accounting for the complex nature of education, (e.g. Blatchford, Bassett, Goldstein & Martin, 2003).

One study that took more account of the complex nature of education, measured several aspects of the Swiss public school classroom environment, including student perceptions of teaching pace and classroom atmosphere (Brühwiler & Blatchford, 2011). The sample was 898 primary and secondary school students and their teachers in the German language region of Switzerland. Primary school students were aged 11-12 years and secondary school students were aged 14 - 15 years. Primary school class sizes ranged between 9 and 24 students with a mean of 18.73 (*SD* = 3.94). Secondary school classes ranged from 14 to 27 with a mean of 20.39 (*SD* = 3.55). Potential effects were measured by assessing a specific taught unit with a pre and post-test measure. Pre-test scores on the measure were standardised on a mean of 50 and a standard deviation of 10. The pre and post-test mean achievement increase was 15.2. Fine-grained aspects of teaching were also assessed as a measure of teacher quality, such as teacher performance and lesson planning. The study found a very small significant effect of a smaller class on primary school students' science learning achievement whereby a one-student decrease in class size equalled 0.5 points in achievement gain. This effect was independent of teacher quality and student/class characteristics such as prior achievement. The effect was greater for classes with more non-German speakers and students with more science knowledge. One limitation of this study is that no clear definitions of class size were given, only the mean and range of class sizes across the conditions were provided. Another limitation concerns the allocation of students to classes. It is unclear whether students and teachers were randomly assigned to different classrooms, or whether they used existing classrooms within the schools. The mean class sizes were also very close to average class sizes of actual Swiss primary and secondary schools at 19.3 and 18.8, respectively. Further, the authors stress that classroom processes measured here did not mediate class size effects. Rather other aspects of a small classroom, untested here, such as increased student attention and effort may also contribute.

A naturalistic design

Increased student attention was examined by Blatchford, Bassett and Brown's (2011) study which used observation methods to assess student-teacher interactions and classroom engagement. The sample included 686 students from UK primary and secondary schools, aged 5-8 years (primary) and11-15 years (secondary). The students were categorized into low, medium and high attaining groups on the basis of teacher ratings. These ratings were also used as a continuous measure. Teacher-student interaction was higher in smaller classes for both primary and secondary classes, with odds ratios of improved interactions of 0.72 for primary school classes and 0.73 for secondary school classes. These results showed an effect of class size on student-teacher interaction but there was no evidence of an interaction between student-teacher relations and attainment. Smaller class size had a positive effect on classroom engagement (odds ratios of 0.73 – 1.12). This was especially so for lowattainers as there was less scope for distraction in a small class. The effect also extended to secondary school. However, it is unclear how class size was defined for each group, as there was no clear distinction given in the study between class sizes.

Classroom Composition And Streaming

Classroom composition, including whether students are streamed/tracked for ability and gender, is another factor that may influence achievement. Classroom composition variables, if not controlled for, are likely to bias class size investigations (Bosworth, 2014). For example, large achievement gains attributed to teacher/class effects may in reality result from ability streaming, whereby more able students are selected for a particular class. Any influence from classroom composition is likely to be small in effect, as teachers do not appear to change their instructional practice in accordance with a change of classroom peer group. (Hattie, 2002). It is often suggested that separating students by gender improves achievement, especially during secondary education where the opposite sex is more likely to be a distraction for adolescent students. Anecdotally, teachers report that behaviour management is easier when dealing with one gender type. This suggests that quality of instruction might be improved in same-sex classes if teachers spend less time on behaviour issues, and in turn may lead to higher achievement. This has indeed been shown whereby the effect of a larger proportion of female peers in a class on higher achievement was mediated by lower levels of disruption, enhanced student-teacher relationships, and reduced teacher fatigue (Lavy & Schlosser, 2011). Partial support was also found for a larger number of female students in the class for South Korean schools (Lee, Turner &

Kim, 2014). This study investigated gender composition in Seoul middle schools where students have no choice in schools they attend. As part of an equalization policy, South Korean students attend their local middle schools without any selection processes. The only selection is for classroom in relation to prior ability. Some schools are single-sex, others have single-sex classes within co-educational (co-ed) schools and others have mixed gender classes within co-ed schools. This set-up removes many of the selection effects that are evident in other education systems. The study found that males in single sex schools did better (0.15 SD) than males in mixed gender classes. Males in mixed gender classes in co-ed schools did better (0.10) than males in single sex classes in a co-ed school. The study found no significant effects of gender composition for female students whose achievement was consistently higher than their male counterparts. The study also suggests that any effect of single-sex school on male achievement is largely driven by increased effort and study time reported by this group compared with the other groups rather than classroom composition. The effect sizes are small between the two types of co-ed schools but the effect is much larger between the two types of single-sex classes (0.21 to 0.28 SD). The study also suggests that teachers in single-sex schools are able to develop specialized strategies to deal with classroom disruption and therefore offers support for such segregation. However, one factor that might affect interpretation of these results is that the single sex schools are more likely to be privately funded compared to the co-ed schools. The studies reported here suggest that reduced classroom disruption, enhanced student-teacher relations, greater academic focus and increased study time mediate the relationship between classroom composition and increased achievement.

Streaming or tracking students in relation to ability has also been suggested to affect student achievement. Such tracking appears to increase achievement for gifted students but has detrimental effects on less able students who are not only likely to feel stigmatized but also become set on a track for life-long low expectations (Hattie, 2002). This is known as the 'Matthew Effect', after the biblical analogy of the rich becoming richer and the poor becoming poorer, whereby individual differences in educational achievement follow this same trajectory (e.g. Stanovich, 1986). The effect has been demonstrated in the German education system where students are tracked into one of three different life courses at the start of secondary education, with almost no chance to change track along the way (Maaz, Trautwein, Ludtke & Baumert, 2008). The German system uses between-school tracking where students are assigned to a low, middle or high track school at secondary level on the basis of prior achievement. There is currently no standardised testing to enter secondary education. Transition into a given track is decided by the primary class teacher who use their own classroom as a frame of reference for prior achievement. Parents may also try to influence the teacher's decision. The three types of school are graded in level of cognitive demand and so if assigned to the lowest track there is little scope for changing track if ability improves later on. Career paths are therefore established at an early age as high track students have an unlimited opportunity to pursue any form of occupation, whereas middle track students are limited to skilled/technical occupations, and students assigned to the lowest track are limited to vocational occupations. A review of the school system found differences between the three trajectories in maths achievement of 0.25 to 0.79 of 1SD (Maaz et al., 2008). It also suggests that strong links are formed between SES and achievement and students with better-educated parents have a higher chance of enrolment into higher track schools. The review suggests standardised testing may go some way to eliminate these effects by selecting students on the basis of cognitive ability rather than classroom position.

Another study that demonstrated the Matthew effect also assessed the three school types in the German high school system. The study found differences in maths achievement between school tracks that increased with advancing grade level (Murayama, Pekrun, Lichtenfeld & Hofe, 2012). This was indicated by a positive association between school type and the total amount of change. The study further tested the relationship between school tracking and the Matthew Effect and found that, when controlling for school track, the significant relationship (r = .29, p =.01) between initial achievement score and total amount of change disappeared (r = .01, p = .79). This suggested that the observed Matthew effect was linked to the school tracking system used in the German education system. The study considers tracking between schools but detrimental effects can also be seen in other education systems that track students between classes within schools.

One study that explored within school effects investigated the influence of tracking on teachers' expectations of student future college attendance (Kelly & Carbonara, 2012). The research was part of the National Education Longitudinal Study (NELS:88) where students and their parents and teachers were followed from junior high school (in 1988) through to employment. The study focused on students who were tracked differently for specific subjects, i.e., high for one subject and low in another. The study was able to control for student background achievement as well as student expectations, as these were collected prior to tracking when in their 8th grade. Student achievement and grades were also assessed along with teacher and student reports of student engagement. Teachers additionally provided data on their own experience and demographics such as qualification and ethnicity. These criteria were also matched with students, e.g. same/different gender. Logistic regression analyses were used to examine potential effects of tracking on teacher expectations. Multilevel logistic regression models were used to explore potential effects of within-student differences in tracking on teacher expectations.

The results showed that most students expected to attend college, with a strong effect of track level on both teacher and student expectations. Students had higher expectations than their teachers for their future college attendance, especially those in low track classes. The study found that grades, test scores, engagement, student expectations and background such as SES shaped teacher expectations; however, tracking predicted beyond these variables. Poor representation of low SES, black, and male students was evident in the high-track courses. Differences in teacher expectation of college attendance were found between the three tracks, with odds ratios of .55 for the low track students, .68 for the middle track and .83 for the highest track. Students were more likely to be expected to attend college by their high track teacher than by their low track teacher. The odds of a positive response by their middle track teachers for future college attendance had a factor increase of 1.29, and for their high track teachers, a factor increase of 1.54. The study showed that student behaviour also differed between classes. For example, student levels of engagement were consistent with strong students in the high track classes and weak students in the low track classes. This may be due to students adapting their behaviour to that of the specific classroom norms associating more closely with peers of the same academic status as themselves (McFarland, 2001).

As the study concludes that tracking influences teacher expectations of college attendance beyond student achievement, this suggests that teachers are biased by the tracking system. However, it may merely be a reflection of achievement and test scores associated with different ability classes, i.e., student achievement or ability is at a level consistent with college attendance for some subjects but not with college attendance for other subjects. The study does not demonstrate whether teachers take account of whole school achievement when making college attendance predictions or whether they are just viewing the student from that specific class or achievement level. If teachers' expectations of student achievement are influenced negatively by ability tracking, it implies that grouping students in this way may be detrimental to their long-term outcomes.

Streaming and academic self-concept

Another aspect of tracking/streaming is its effect on academic self-concept. Academic self-concept is one's own evaluation of their academic performance or ability based on student experience (Shavelson, Hubner, & Stanton, 1976). Research has shown that self-concept and achievement are reciprocally linked, leading to increases in each other (e.g. Marsh & Craven, 2006). Individuals make comparisons between themselves and a social frame of reference to inform their self-concept, which in the case of academic self-concept will be the school or classroom (e.g. Bong & Skaalvik, 2003). This has led to extensive research exploring the relationship between achievement and academic self-concept, with some researchers suggesting that being grouped with peers of high ability, can lead to reduced academic self-concept despite being equally able (Marsh, 1987; Trautwein, Lüdtke, Marsh, Köller, & Baumert, 2006). Other research has shown that students' academic self-concept depends on the type of tracking they are subject to (Chmielewski, Dumont, & Trautwein, 2013). Students in schools where one or two specific courses were tracked demonstrated higher levels of mathematical self-concept if they were in the high track, and lower mathematical self-concept if in the lower track. The study suggested that daily regrouping of students continually expanded the students' frame of reference to include their whole year group, and so they regularly placed themselves at extreme ends of their school. When compared to the whole year group, students in the top track would repeatedly feel higher levels of self-concept and those in the lower track would repeatedly feel lower levels. Whereas being grouped with peers across all subjects, as in fully tracked or mixed ability schools, students' frame of reference would be much smaller: only confined to their immediate peers (Chmielewski et al., 2013). The research suggests that academic self-concept is a potential mediator between school tracking and achievement (e.g. Marsh & Craven, 2006).

Teacher Characteristics

Teacher characteristics are often overlooked in educational research despite their potential role in classroom effects. One obvious characteristic is teacher qualification, and more specifically qualification in the subject being taught. Research that investigated this found significant differences in maths achievement scores between students whose teachers had a maths degree and students whose teachers did not (Rowan, Chiang, & Miller, 1997). The study was conducted using the NELS:88 sample of 5,381 10th grade students (aged 15 to 16 years) and their teachers. However, the effect of 0.015 SD seen here is small. Greater effects were found for teachers' mathematical knowledge in a study of 181 German high school teachers of 4,353 15-16 year old students (Baumert et al., 2010). The study found that pedagogical content knowledge (PCK) had greater effect on achievement explaining 54% (R^2 = .54) of between-class variance, compared to content knowledge (CK), which explained 44% (R^2 = .44). A difference in teachers' PCK of 2 SD led to a difference in their students' maths achievement by *d*= 0.46 SD across tracks and *d*= 0.33 SD within tracks. The study controlled for several factors: student maths and literacy achievement, mental ability, parental education, social status and immigration status; teachers' own final grade point average, effective classroom management, and instructional quality. Maths teachers, however, are trained differently depending on which track they will be teaching. For example, they are taught more advanced content if they intend to teach the highest ability track. This was borne out by the significant difference in CK between the highest track teachers and the other tracks. The study was unable to distinguish whether the higher CK was due to higher demands of teacher training in the higher academic track and/or higher demands of school maths departments in the higher track schools. The study reported that when controlling for CK, middle track teachers had higher PCK than the other tracks. Students in the lowest tracks appeared to have teachers with lower levels of CK and PCK. Differences in curricular and teacher training are confounded with CK and PCK.

Another aspect of teacher competence is instructional practice. This was investigated recently using data from the Trends in International Mathematics and Science Study (TIMSS, 2007; O'Dwyer, Wang, & Shields, 2015). Maths achievement in relation to teachers' instructional practice was assessed in a sample of 12,346 students aged 13 to 14 years from the US, Korea, Japan and Singapore. Differences between countries were found in their use of six instructional practices, which included '*apply facts, concepts and procedures to solve routine problems';* '*work on problems for which there is no immediately obvious method of solution'* and *'write equations and functions to represent relationships and interpret data in tables, charts or graphs'.* US teachers reported their use of four out of six of the practices more frequently than teachers from other countries. The US teachers were less likely to use the other two practices than Japan and Korea. Although initial associations were found between instructional practice items and maths achievement for US and Singapore students, these explained very little of the variance in maths achievement after controlling for gender; home background; positive affect towards maths; valuing maths; and maths self-confidence (between 0.2 and 1.6 additional percentage points). No associations were found for Korea and Japan between achievement and the six practices. The study makes the point that in many East Asian countries extra-curricular tutoring is widespread, and therefore it is difficult to make between-country comparisons in teacher instruction and maths achievement in schools without also accounting for this variable (Bray & Kwo, 2013).

The previous studies focus on teachers' ability to teach the subject with regard to subject knowledge and instructional technique, however, related to this is teachers' confidence in their ability to teach the given subject. Research has shown a positive association between teacher self-efficacy and achievement in both maths and reading (Goddard, Hoy & Woolfolk-Hoy, 2000). The study conducted in a sample of 7,016 US students aged 7 to 11 years (grades 2, 3 and 5), showed that collective teacher self-efficacy explained 53.3% of the between-school variance for maths achievement and 69.6% of the between-school variance for reading achievement. However, school selection for ability may also influence these results.

Teacher self-efficacy encompasses more than just confidence in ability to teach the subject at an instructional level, it also extends to the ability to engage and motivate students as well as classroom management (Tschannen-Moran & Woolfolk-Hoy, 2001). More recently, a study that investigated teacher self-efficacy in a sample of 2,184 Italian high school teachers discovered a relationship between self-efficacy, job satisfaction and student achievement (Caprara, Barbaranelli, Steca & Malone, 2006). The study showed a small (1%, R^2 =.01) significant effect of student achievement at time one on teacher self-efficacy at time two. However, the effect is very small, in comparison to the direct effect found of teacher self-efficacy at time 2 on student achievement at time 3 of 8.2%(R^2=.082). In this study, no influence of teacher job-satisfaction on student achievement was found. One limitation of the study was the inability to detect any potential reciprocal effects of self-efficacy and achievement.

Another suggested effect of teacher/class on achievement is the ability to create an optimum emotional environment within the classroom. One study investigated emotional quality, instructional quality, and subject exposure within two domains: reading and maths (Pianta, Belsky, Vandergrift, Houts, &

Morrison, 2008). The longitudinal study used data from the National Institute of Child Health and Development study (NICHD) of 1,364 US children aged 54 months at the study commencement across grades 1, 3 and 5. Two achievement trajectories were found for reading, fast and typical, and one trajectory for maths. For the fast reading trajectory, students' skills improved quickly before levelling. The typical reading trajectory students' skills grew less rapidly than the fast group. Interestingly, the fast trajectory of reading skills was not influenced by emotional quality, instructional quality or literacy exposure. However, for the typical trajectory the study reported a positive effect of higher emotional quality on reading scores with 1 point above the mean leading to a 1.6-point increase at 3rd grade; and a 3.7-point increase at 5th grade. An interaction was also reported between emotional quality and quantity of exposure at 1st grade with less improvement shown for the low emotional quality classrooms (0.13 points). For maths, only emotional quality influenced at grade 5 with scores 1 point above the mean leading to a 2.4-point increase in maths scores. Quantity of instruction had some influence with 1 point above the mean leading to an increase in maths score of 0.28 points at 3rd grade and 0.35 points at 5th grade. However, quality of instruction was not reported to moderate the quantity of exposure to maths instruction. The effects reported here appear to be extremely modest. For example, when considering the mean score at 3rd grade of 494.33 (*SD*=15.84), a 1.6 point increase seems negligible. The study did not provide sufficient information on what the point increase referred to (raw or standardised scores) therefore it is not possible to evaluate effect sizes. Further it was a field study of individual target children at different schools rather than multiple children in each classroom, which would have enabled multilevel modelling and greater statistical confidence in results.

The previous studies demonstrate how certain teacher characteristics have a positive influence on student achievement. Research has also shown a negative relationship between student math achievement and teachers' own maths anxiety (Beilock, Gunderson, Ramirez, & Levine, 2010). The study investigated this relationship in a sample of 117 US 1st and 2nd grade students (aged 6 to 8 years across the grades) and their 17 female teachers. Data were collected for student maths achievement and gender ability beliefs at two assessment points: at the beginning and at the end of the academic year. At the end of the year teacher maths anxiety and maths knowledge were assessed. The study found that girls with traditional gender ability views (i.e. boys are good at maths and girls are good at reading) had significantly lower maths achievement than girls without these views (d =0.66), and boys (d =0.37). The study also found that higher teacher maths anxiety associated with traditional gender ability beliefs in girls only (r = 0.28, p<.05). Traditional gender beliefs also negatively associated with time 2 maths achievement for girls only (r = -0.28, p<.05). Teacher maths anxiety (β = −0.21, t = −2.17, p = 0.034) and traditional gender ability beliefs (β = -0.23, p<.01) predicted lower maths achievement at time 2 for girls only. The relationship between teacher maths anxiety and girls' maths achievement was mediated by traditional gender beliefs as this relationship disappeared when gender beliefs were also included in the model. While the associations shown between teacher maths anxiety, gender ability beliefs, and maths achievement were significant, the associations were modest. Additionally, although teacher knowledge was tested, the study does not include how this may also have influenced relationships. Equally, it would have been interesting to know the role of student maths anxiety in these relationships; however, this was not assessed. Furthermore, the first assessment was taken within 3 months of start of term but reported as start of academic year. This may not have provided an adequate baseline; greater effects may have been observed if data were collected earlier in the academic year.

Behavioural Genetics Research

The studies reviewed in the previous section show small effects on student achievement under the influence of the teacher/class. The potential sources of influence are interrelated and/or mediated by other variables, so that the actual causes and effects are difficult to establish. Moreover, the most obvious potential confound – individual variation in student ability – is often not controlled for. Behavioural genetic research methods can help with establishing causal paths. Many genetically informative studies have evaluated the sources of the individual differences in ability and achievement, as well as the sources of associations between hypothesised effect and outcome variables, For example, one UK study investigated relationships between science performance and the science-learning environment in a sample of 3000 14 year old twin pairs from the Twins Early Development Study (TEDS; Haworth, Davis, Hanscombe, Kovas, Dale & Plomin, 2013). The twins' experience of their learning environment was obtained by self-report and included their perceptions of peer environment and teacher-student interactions. An online test of scientific enquiry was also completed to assess science performance. Behavioural analyses found some gender differences with boys perceiving a more positive peer environment than girls. Boys also demonstrated enhanced science performance; however, the effect sizes were very modest with 0.40% of the variance explained by peer environment and just 0.20% of the variance explained by science performance. The results showed moderate heritability (43%) for the science-learning environment, with moderate non-shared environment (54%) supplying the larger part. Very little contribution was made from the shared environment (3%). Bivariate analyses showed a modest genetic correlation (r=0.27) between learning environment and science performance, indicating that some of the genetic factors that influence science learning environment also influence science performance. Separate analyses showed that learning and peer environments were very similar, with a substantial genetic overlap (98%) between them that suggests, in terms of heritability, peers are an important part of the classroom environment. The heritability of the peer environment indicates a gene-environment correlation whereby an individual establishes or seeks out environments that are associated with their own genetic propensity (e.g. Haworth, Asbury, Dale, & Plomin, 2011). An individual's genetic propensity will influence to some extent how they respond to their peers and how they select them. Students do not passively experience the learning environment; to a great extent they will evoke responses from their environment (Plomin & Bergman, 1991). Their genetic propensity will influence their teacher-student interactions and in turn impact their learning environment. The peer and learning environments were measured by self-report, which asked about their perceptions of teacher-student interactions within the class and for peer perceptions focused mainly on interactions about science outside of the classroom. This focus outside the classroom implies students had more choice in these interactions and therefore allowed more opportunity for genes to play a role.

Beyond estimating genetic and environmental sources of variance and co-variance in educational traits, twin studies can provide another piece of information relevant to classroom effects. It is frequently suggested that any differences between twin pairs (especially identical) taught separately must be due to the classroom environment and should therefore constitute a teacher/class effect. One study, using the TEDS sample, found that twins in separate classrooms were only marginally more different in school achievement and cognitive abilities such as verbal and non-verbal reasoning, at ages 7, 9, and 10 years in comparison to those taught together (Kovas, Haworth, Dale, & Plomin, 2007). Further, twins in different classrooms were no more different in their academic motivation than twins in the same classrooms at age 9 (Kovas et al., 2015). These studies suggest that any influence of the classroom is modest beyond other student factors.

Another investigation into twin pair similarity, explored potential classroom effects in association with reading ability (Byrne, Coventry, Olson, Wadsworth, Samuellson, Petrill, Willcutt & Corley, 2010). The study examined the suggestion that variation in teacher characteristics is a principal contributor towards differences in pre-schoolers' early literacy achievement. The study found, in two samples of Australian and US preschool students, higher correlations in literacy between twin pairs taught within the same classroom, compared with twin pairs taught separately. Although the majority of analyses did not reach significance, the results generally showed higher correlations for monozygotic (MZ) and dizygotic (DZ) twin pairs in the same class compared to twin pairs taught separately. This suggests slightly greater similarity in twins taught together than twins taught separately. Higher correlations were also shown for monozygotic (MZ) twins compared with dizygotic (DZ) twins across time suggesting a genetic contribution towards stability in literacy development. These findings are also borne out by greater mean differences shown for DZ compared to MZ twins (.096 and .066 respectively). The longitudinal study showed that the pattern of results existed since kindergarten. However, it was demonstrated that twin pairs were not assigned to separate classrooms as a consequence of pre-existing literacy differences between them. Additionally, MZ twin pairs moving from the same to different classrooms did not deviate in growth as would be expected if classroom effects were robust. As average differences were shown to be .08, Byrne and colleagues estimated the variance ascribed to ‘teacher effects’ as 8%, and suggested that actual teacher characteristics would contribute towards this figure in combination with other classroom influences, including curriculum.

Similar findings were revealed by Taylor, Roehrig, Soden Hensler, Connor, and Schatschneider (2010) when investigating teacher quality in relation to genetic and environmental variance in seven year old students’ reading achievement. Teacher quality was measured by residualised growth in the twins’ classmates’ oral reading fluency scores while controlling for their initial levels. The results showed that the unique heritability in reading scores was moderated by teacher quality: heritability of reading was greater for students with teachers of higher levels of quality. These results suggest that genetic influences on achievement may be moderated, to a small extent (5%) by teacher quality. An additional study within the same sample conducted by Hart, Logan, Soden-Hensler, Kershaw, Taylor and Schatschneider (2013) also found small effects in growth between classes in reading, emphasizing the conclusion that influences of teacher quality on student achievement are small.

The previous studies, while showing some effects of the classroom on reading outcomes, use growth in student performance to index teacher/classroom effects. In fact, this has been the case across the literature. Growth in student achievement however is an outcome which can be accredited to several factors not quantified within these studies (Olson, Keenan, Byrne, & Samuelsson, 2014). Ignoring these other factors leads to credit being misdirected towards just one component of the learning environment. While teacher characteristics are likely to be important for optimum learning, other factors also contribute towards an individual's learning environment to a greater or lesser extent. These may be further affected by individual differences in motivation, attitudes towards and interest in the subject as well as students' peers (Marsh, Martin & Cheng, 2008). This raises the issue of how much effect the class/teacher may have on non-cognitive factors like motivation. Research has shown that shared environment, such as class or home factors that are expected to contribute towards similarities among family members, did not contribute towards similarity in twins' motivation. Instead, genetic factors explained 40% of the variance in motivation with the remaining 60% accounted for by non-shared environmental factors (Kovas et al., 2015). This shows that a substantial part of student achievement stems from individual-specific factors. This is not to say that class effects are negligible, rather, they may be perceived differently by individual students and therefore are not 'class-wide'.

One study has indeed shown that class effects were perceived differently. Asbury, Almeida, Hibel, Haarlar and Plomin (2008) investigated perceptions of classroom experience in relation to English, maths and science achievement in a sample of 121 monozygotic (MZ) twin pairs. The twin pairs, who were recruited from the Twins Early Development Study (TEDS), were taught within the same classrooms and so shared their teacher and peers. One might assume that, as MZ twins share 100% of their segregating genes, their perceptions would be highly similar, however, this was not the case. A design was used to investigate MZ differences in classroom perceptions obtained by telephone interviews conducted every day for a two-week period. MZ differences in maths achievement associated negatively and significantly with MZ differences in peer problems, with an effect size of 8%. A significant association was also shown between MZ differences in school positivity and differences in maths achievement, with an effect size of 15%. Additionally, an effect size of 8% was revealed for differences in school positivity and differences in science achievement. This study suggests that students have different perceptions of the same classroom experience and these differences are relevant to differences in academic achievement. Furthermore, the effects appear to be for maths and science as opposed to English, which did not yield the same significant associations.

Conclusion

Overall, the literature suggests small teacher/classroom effects on students' academic motivation and achievement. The large-scale survey studies, while having access to an enormous amount of data, are deficient in their ability to detect effects due to methodological issues such as questionable measurement of teacher quality/VA. Research that examined class size found modest effects on achievement for higher (but not lower) achievers. However, it was unclear in many studies what constituted small or large classes. Classroom composition seems to have some effect for male students but not for female students. Some small effects from between-school tracking were seen in the German education system, although other factors apart from SES need to be considered. Research into within-school tracking shows how teachers can influence students but are also subject to school effects themselves. Teacher ability also has influence in terms of subject knowledge and self-efficacy. Again though the effects are modest. Any effects seen may be due to ability selection processes, which can be implicit or explicit. It appears from the literature that schools, teachers and classrooms have some effect but other factors, such as student specific characteristics and perceptions, play the largest role. Although much research has been conducted, very few conclusions regarding teacher effects are solid. This does not suggest that schools and teachers are unimportant. On the contrary, schools and teachers are very important as without them children would not be acquiring the curriculum that has been developed as a necessary body of skills and knowledge for functioning in modern society. However, the differences in motivation and learning seem to largely stem from individual specific characteristics, rather than class-wide effects.

Using unique pseudo-experimental methods, the present thesis aims to address these issues by investigating teacher/classroom effects on motivation, performance, and school achievement while taking account of several aspects of the classroom environment. Employing a longitudinal design, these factors are explored to see how they unravel across several assessments during one academic year. The investigation also uses a cross-cultural approach, which allows the comparison between two different education systems, in Russia and the UK (Chapters 3, 4, 5 and 6). In so doing, the thesis is able to take account of the differences in streaming and tracking processes between the education systems. Tracking and streaming processes are employed in the UK, but they are not formally applied in Russia. The thesis investigates students in their first year of secondary education where they have specific subject teachers for the first time. This also allows the study of two different domains to investigate different classrooms and teachers for the same students. Maths and geography classrooms are compared as both domains contain mathematical content but the material is taught and perceived very differently across the two subjects. In this investigation, actual school achievement as graded by the class teacher is used as a more reliable outcome measure relevant to the curriculum. Performance is also assessed independently to provide a more objective outcome measure. The investigation also takes account of within-classroom factors, which include, student-teacher, peer-peer relations, and the calm/chaotic atmosphere of the classroom. In addition to these factors, teacher characteristics are included to assess experience, emotional ability and self-efficacy in teaching and classroom management. Student characteristics are considered through an assessment of motivational factors and subject anxiety. Such factors provide a more fine-grained approach to investigate teacher/classroom effects.

Using another pseudo-experimental approach to investigate teacher/classroom effects, similarities and differences are also investigated between twin pairs taught together, and twin pairs taught separately (Chapter 7). Students from two large twin samples from the UK and Quebec (Canada) are followed longitudinally from ages 7 to 16 years and assessed on measures of school achievement, motivation and cognitive ability. Differences in education systems are also considered, as the education system in Quebec does not formally apply streaming and tracking processes, unlike the UK. Greater differences between twins if taught separately over those taught together could imply an effect of teacher/classroom. This approach, together with the fine-grained approach, provides a more comprehensive investigation into teacher/classroom effects.

The Aims Of The Present Thesis

This thesis sets out to investigate:

1. Whether there are differences between the Russian and UK samples in academic outcomes, i.e. performance, classroom environment, motivation, subject anxiety, classroom atmosphere, homework behaviour and feedback, attitudes towards the subject (Chapter 3).
2. Whether potential differences in these constructs persist across the academic year (Chapter 3).
3. Whether the patterns of results are similar for maths and geography (Chapter 3).
4. Whether being among the same peers for the previous four years and continuing, has an overriding influence beyond the class teacher (Chapter 4).
5. Whether having the same primary school teacher for the previous four years influences the classroom environment beyond the current subject teacher (Chapter 4).
6. Whether teacher/classroom effects are similar across different domains, i.e. maths and geography (Chapter 4).
7. Whether significant effects of classroom and teacher groups found at time 1 persist across time 2 and time 3 (Chapter 5).

8. Whether patterns of class rankings found at time 1 are also maintained across subsequent waves (Chapter 5).
9. Whether potential patterns of significant effects and rankings persist when taking account of prior achievement (Chapter 5).
10. Whether potential significant effects and ranking patterns persist in the same way across maths and geography class and teacher groups at time 2 and time 3 (Chapter 5).
11. Whether potential significant effects and ranking patterns found in the Russian sample are similar to any potential effects found in the UK sample (Chapter 5).
12. Whether teacher characteristics in the Russian sample, associate with classroom environment measures and performance/achievement (Chapter 5).
13. Whether teacher characteristics in the Russian sample, mediate potential relationships between classroom environment measures and performance/achievement (Chapter 5).
14. Whether associations between maths anxiety and maths performance develop differently for students in Russia and the UK (Chapter 6).
15. Whether reciprocal associations exist between geography anxiety and geography performance, as previously shown for mathematics and other academic domains (Chapter 6).
16. Whether associations between geography anxiety and geography performance develop differently for students in Russia and the UK (Chapter 6).
17. Whether there are average differences in school achievement, cognitive ability and motivation between twin pairs taught together (i.e. by the

same teacher/class) and twin pairs taught separately (i.e. by different teachers/classes) (Chapter 7).

18. Whether there are any differences in separation effects, in light of differences in timing of separation, purpose of separation (e.g. streaming; policy recommendations) and twins' sex or zygosity (Chapter 7).

Chapter 2 Pilot study

Chapter 3 investigates potential differences and similarities between Russia and the UK across one academic year on measures of test performance, classroom environment, motivation, subject anxiety, classroom atmosphere, homework behaviour and feedback, and attitudes towards the subject within two domains, maths and geography. Potential differences are also assessed for perceptions of intelligence and socioeconomic status. The study uses data collected longitudinally across three assessment points over the course of one academic year in four urban schools, two in the UK and two in Russia. All schools are mixed ability, although in the UK, students are streamed by ability for their maths classes. In Russia, the students attend a school where they have the opportunity to learn two second languages.

Chapter 4 investigates potential teacher/classroom effects in a sample of Russian 10-12 year old students who are not streamed for ability and remain in the same class groups throughout their school education. Using a cross-sectional approach to investigate at one assessment point, the study explores classroom and teacher differences across the range of measures investigated in Chapter 3 with the addition of school achievement. Potential differences are investigated across two domains, maths and geography.

Chapter 5 extends from **Chapter 4**, and investigates longitudinally, the continuity of effects for maths and geography classrooms and teachers. In particular, the study aims to investigate whether potential significant effects explored at time 1 (the first assessment wave in January) persist across the academic year at time 2 (April/May) and time 3 (September, following the summer break). In addition, this study also explores whether similar effects are found in the UK sample. The study also investigates potential associations between teacher characteristics and measures of classroom environment and performance.

Chapter 6 investigates the development of associations between academic anxiety and academic performance, comparing across two countries, Russia and the UK, and two domains, maths and geography. The study uses a longitudinal and cross-cultural design to explore whether academic anxiety and academic performance develop differently across the two samples, given the differences in education systems between the two countries.

Chapter 7 investigates teacher/classroom effects in relation to twin pairs taught together or separately. The study investigates longitudinally within two large twin samples from the UK and Quebec (Canada), whose participants were followed longitudinally from ages 7 to 16 years and assessed on measures of school achievement, motivation and cognitive ability. Differences in education systems were also considered as in the UK, students are streamed by ability for their maths classes, whereas students in Quebec are not. Greater differences between twins if taught separately over those taught together could imply an effect of teacher/classroom.

Chapter 2

Measuring teacher/classroom effects on educational outcomes: pilot study

Introduction

Extensive planning took place prior to commencement of the research, including selection, translation and adaptation of measures. The biggest challenge to planning of the study was selection of specific aspects of the school environment for investigation. Attempting to capture every aspect of student/teacher/class would require such extensive data collection to render the study impractical. In order to minimize the load on participants and disruption to the schools, key measures were selected to enable participation within the 55 minutes of one lesson period.

For the student measures, the study mainly selected those that had been previously used in the Twins Early Development Study (TEDS; Haworth, Davis, & Plomin, 2013), as these are well-established, reputable instruments used extensively in the literature and additionally validated in the TEDS research. As some measures had been devised for different ages to the current sample of 11 to 12 year old students, some adaptation in line with their stage of curriculum was needed. It was also necessary to extract the highest amount of information using the least number of items; so further adaptation was necessary to avoid any overlapping items. Furthermore, the instruments had mostly been developed to assess mathematics and so additional modifications were required to assess geography.

Teacher measures were selected by reviewing the previous literature on education and occupational research. The instruments were selected to investigate six domains of teaching: demographics, experience, job-satisfaction, self-efficacy, emotional intelligence and occupational burnout. Teachers' perceptions of classroom atmosphere were also assessed.

The Pilot data collection was conducted in the UK during July 2013 with the aim to test the timing and validity of the revised measures for students and feasibility of measures for teachers.

Methods

Ethical approval

Ethical approval for the study was provided by Goldsmiths' Ethics Committee in June 2013, prior to data collection for the pilot study. As participants were under eighteen years, it was necessary to obtain consent from their parents/guardians. Approval was granted to obtain parental consent via an opt-out process. Any parents/guardians who did not wish their child to participate were given the option to exclude their child from the study by returning a completed opt-out consent form that was sent out to all students' parents/guardians.

Participants

A sample of 38 (19 male, 18 female, 1 missing) 11-12 year old (*M*= 149 months (*SD*= 3.71)) year 7 students and two teachers from one of the UK participating schools took part. The students were of the same age and grade (year 7) as the UK and Russian sample used in the main study, but were from the previous year's student cohort.

Measures

The measures are described here in detail. Appendix 1 presents results of analyses to test internal reliability of the measures for the Russian and the UK samples whose data were used in Chapters 3, 4, 5, and 6. Descriptive statistics for all assessed variables are also presented in Appendix 1 (Tables 1.6 to 1.8 present student measures and Table 1.9. present teacher measures). The majority of measures demonstrated normality for both samples at each assessment. Measures that did not meet normality at some assessment waves were geography performance which was negatively skewed in the UK sample, number line which was positively skewed in the Russian and UK samples, and homework behaviour for both geography and maths were negatively skewed in the Russian and UK samples. These variables were transformed and used throughout the thesis where direct comparisons were made across assessments and samples.

Student Cognitive Measures

Maths performance. This was tested with the Maths Problem Verification Task ((M)PVT; Murphy & Mazzocco, 2008). A mathematical equation was presented and the participants had to indicate by placing a cross in the appropriate box whether the equation was right or wrong; there was an option if they did not know. The whole task (48 items) was timed and the participants had 8 minutes to complete the test. Each correct item was given a score of 1 and any other response, including 'don't know', was given a score of

0. The total score was the sum of scores for the 48 items, giving a total score out of 48. Example items include: 13x4=47 (wrong), 2/6= 3/9 (right).

Number estimation. This was assessed with The Number Line Task (Siegler & Opfer, 2006). This task has two variants: the Number-to-Position (NP) Task and the Position-to-Number (PN) task. This study used the NP task, whereby participants were shown a number and asked to estimate its position on the number line that ranges from 0 to 1000. The measure was scored by converting estimates of linear magnitude into a real number. To do this, the distance was measured from the left end point to the hatch mark (in linear units), that distance was then divided by the total length of the line, and then multiplied by the number given on the other endpoint. The target number was then subtracted from this calculation to provide a 'score'. Successful item estimates gained scores close to zero; inaccurate estimates gained scores far from zero (either positive or negative). The total test score was the mean of the absolute (positive) values of the item scores, rounded to one decimal place. Most test scores were in the range of 10 to 100. The final mean score was recoded to absolute values to remove negative numbers. As some analyses do not compute with zero values, a second version of the variable was computed adding '1' to all mean scores.

Geography performance. This was assessed with the Geography Problem Verification task (GPVT). The measure, developed in collaboration with Russian colleagues, was adapted from the maths problem verification task and uses the same principles. Participants were presented with statements and they had to indicate, by placing a cross in the appropriate box, whether they thought it was right, wrong or they don't know. Statements related to the solar system; directions on a map (north, south, east and west); time zones; and meridian lines. Example statements included, ‘the equator divides the earth into two equal hemispheres’; and ‘on earth there are six oceans’. Participants were presented with two practice items which did not contribute to the total score. Each correct item was given a score of 1 and any other response, including ‘don’t know’, was given a score of 0. The total test score was the sum of the item scores for the 37 main test items, giving a total score out of 37.

Student Maths-Related Non-Cognitive Measures

Subjective measures of enjoyment and ability for maths. These were assessed with Self-perceived Ability and Enjoyment for Maths (Spinath, Spinath, Harlaar, & Plomin, 2006). Taken from the Twins Early Development Study (TEDS) booklet used when the children were 12 years of age, the measure comprised of two separate questionnaires of three items each, one asking ‘*How much do you like…?’* and the other asking ‘*How good do you think you are at…?’* The original measure assessed perceived ability and enjoyment in all academic subjects but this study used the six questions that related specifically to mathematics. Using a 5-point scale, perceived ability ranged from ‘very good’ to ‘not at all good’. Enjoyment ranged from, ‘like it very much’ to ‘don’t like it at all’. Example questions included, ‘*solving money and number problems*, and *doing maths in your head’*. The items were scored 1-5, providing a mean score out of fifteen. Higher scores indicated greater enjoyment and perceived ability.

Subjective measures of classroom environment. These were assessed with Maths Classroom Environment, using 12 items from a 19-item measure taken from ‘Your School’ questionnaires used in the TEDS when the children were 16 years of age. Students were asked to think about their maths classroom environment and teacher in the past year (in this study, since the beginning of term) and rate which statements were true for their classroom. Classroom items included, '*some pupils try to be the first ones finished'*. Teacher items included '*the teacher shows an interest in every student's learning'*. A 4-point scale was used ranging from 0 = 'never' to 3 = 'every lesson'. The original measure was adapted from two questionnaires: Student classroom environment, 9 items adapted from the full 12-item measure (Midgley, Eccles & Feldlaufer, 1991); and 10 items from PISA – classroom environment. This study used 12 of the original 19 items, to avoid items inappropriate for this stage of education and prevent overlap with other measures. In the main study, factor analysis on the measure revealed two subscales: student-teacher relations and peer competition (see Appendix 2 for details). The total scale was assessed in the Pilot Study, in the main study (**Chapter 3, Chapter 4** and **Chapter 5**) the subscales were also used.

Maths homework and feedback. This was assessed with the PISA Maths Homework Questionnaire selected from PISA student questionnaires (2000, 2003 & 2006) from 'Your School' questionnaires used by TEDS 16 year study. The measure assessed participants' attitudes towards homework and their perceptions of teacher feedback. Students were asked to indicate how often each of the five statements relating to homework applied to them based on a 4-point scale ranging from 0 = 'never' to 3 = 'always'. Items included, '*I complete my homework on time'*; '*my teachers make useful comments on my homework'*. The measure divides into two subscales, *Homework Behaviour* and *Homework Feedback.* The total scale was assessed in the Pilot Study, in the main study (**Chapter 3, Chapter 4** and **Chapter 5**) the subscales were also used.

Maths classroom atmosphere. This was assessed by the Classroom Chaos questionnaire. Participants' gave their perception of how calm or otherwise they viewed their classroom. The measure was adapted from the TEDS teacher booklet used when the children were 10 years of age. Classroom CHAOS was originally adapted from the Confusion, Hubbub and Order Scale developed to assess the home environment (Matheny, Wachs, Ludwig and Phillips, 1995). In this study, one item from the teacher's perspective was dropped when administered to students. Participants were asked to rate 'yes' or 'no' for 15 statements, items included: '*My classroom is a place where: there is very little noise*'; and '*we almost always seem to be rushed'*. The measure was scored 1/0 (yes/no); a high score indicated low chaos.

Maths environment. This was assessed with the Maths Environment questionnaire taken from the 15 item *Attitudes to maths and reading* measure used in the TEDS study at age 10 years as part of a student background questionnaire. Just three items were used which were specific to maths and to avoid overlap with items in other measures. These were answered on a 4-point scale: 'Never or hardly ever' = 1 to 'almost every day' = 4. Items included, '*how often do you solve math problems with a partner or in small groups?';* and '*how often do you work with objects like rulers, counting blocks, or stopwatches*?'. Adapted from NAEP (2005).

Maths tutoring. This was developed to assess whether students had any extra tuition outside of the maths classroom environment. The measure comprised of 3 parts: (1) a yes/no question: '*Do you have any extra tutoring for maths outside of school?'.* If answered yes, (2) '*How many hours a week?',* responses on a 5-point scale ranged from 1 to 5 hours; and (3)'*What are the reasons for your extra tuition?'; 'I like maths'; 'I struggle with maths'; 'My parents* *make me attend extra classes'.* Responses ranged from 'strongly agree' to 'strongly disagree'.

Maths anxiety. This was assessed with the Abbreviated Maths Anxiety Scale (AMAS; Hopko, Mahadevan, Bare, & Hunt, 2003); currently the shortest valid maths anxiety measure. It comprised only 9 items and shown to be equally effective as the longer 25 item MARS (Hopko et al., 2003). Participants were presented with a statement and asked to assess how anxious that situation would make them feel on a 5-point scale where 1 = 'not at all' and 5 = 'very strong'. Example items included: '*thinking about an upcoming maths test, one day before'*.

Attitude towards maths. This was assessed with Maths Usefulness (PISA, 2000, 2003 & 2006), which presented participants with four statements regarding their perception of the usefulness of maths. The measure was adapted from the PISA Attitudes Towards School questionnaire used in the TEDS study at age 16. Items included: '*maths classes have been a waste of time'*, and '*maths classes have taught me things which could be useful in a job'*. Participants were asked to rate their agreement with these statements on a 4-point scale ranging from 1 = 'strongly disagree' to 4 = 'strongly agree'.

Student Geography-Related Non-Cognitive Measures

All the above non-cognitive measures were adapted for participants' geography classrooms.

Subjective measures of enjoyment and ability for geography. These were assessed with Self-perceived Ability and Enjoyment for Geography. It was adapted from the mathematics version above and follows the same format asking '*how much do you like…?'* And '*how good do you think you are at…?'.*

Example items included: '*solving direction and route finding problems';* and '*visualizing locations of places in your head'*.

Subjective perceptions of classroom environment. These were assessed with Geography Classroom Environment using the same measure used for the maths classroom environment. Participants were asked to think of their geography classrooms since the beginning of term when answering the items which were generic for all school subjects.

Geography homework and feedback. This was assessed with the same measure used for maths homework (PISA, 2000, 2003 & 2006), which is generic for any subject and so was used asking participants to think of their geography classrooms when responding.

Geography classroom atmosphere. This was assessed with Classroom Chaos (Matheny et al., 1995) as used to for maths classroom atmosphere. The measure is generic to any classroom and so participants were asked to think of their geography classroom when responding.

Geography environment. This was assessed with the same three items used to assess maths environment (NAEP, 2005) but adapted for geography. Items included: '*how often do you solve geography problems with a partner or in small groups?';* and '*how often do you work with objects like rulers, compasses, atlases, or maps?*'.

Geography tutoring. This was assessed with the same measure used for extra maths tuition but adapted for geography.

Geography anxiety. This was assessed with the AMAS (Hopko et al., 2003), used above for maths, and was adapted for geography. Items included: '*having to use compass directions on a map';* and *'watching a teacher work a route finding/direction problem on the whiteboard'*.

Attitude towards geography. This was assessed with Geography Usefulness using an adapted version of the above measure used for maths (PISA, 2000, 2003 & 2006). Items included: '*geography classes have been a waste of time'*.

Student Perceptions Of Intelligence And Socioeconomic Status

Perceptions of intelligence. This was assessed with Theories of Intelligence (Dweck, Chiu, & Hong, 1995). This three item measure assessed individuals' beliefs regarding intelligence, i.e. whether it is fixed or changeable. Participants were presented with three statements and asked to rate on a 6 – point scale how much they agreed or disagreed with them. Items were scored 1 = 'strongly agree' to 6 = 'strongly disagree', and included: '*You have a certain amount of intelligence and you can't really do much to change it';* and '*You can learn new things, but you can't really change your basic intelligence'*. Participants were classified as entity theorists if their overall implicit theory score was 3.0 or below and classified as incremental theorists if their overall score was 4.0 or above.

Perceptions of academic and socioeconomic status. These were assessed with the MacArthur scale of subjective social status (Adler & Stewart, 2007) and asked four questions regarding participants' perceptions of how they fit in at school or college regarding respect from other students, their perceptions of academic ability in relation to other students, how their family fits in with British/Russian society regarding both employment status and education level. They were asked to do this by placing themselves on a 10 – rung ladder which represented their school/college or British/Russian society, where the first rung was the bottom and the tenth was the top.

School Achievement Measures

UK baseline measures of achievement. *SAT Scores* (www.gov.uk) or key stage 2 National Curriculum tests are designed to test UK students' knowledge and understanding of specific elements of the key stage 2 programmes of study. They provide a snapshot of a student's attainment at the end of the key stage. English and mathematics tests are taken by children at the end of year 6, usually at aged 11 years. Level 6 tests form part of the suite of key stage 2 National Curriculum tests. These are optional and are aimed at high attaining children. At the end of key stage 2, teachers assess students' attainment in English, mathematics and science. These teacher assessment judgments are reported to the Standards and Testing Agency (STA) as well as to parents. The results are often used by UK secondary schools for streaming/tracking students in their maths classes. As the UK participants completed these tests at primary school in year 6 prior to commencing secondary school, these data were made available to the study.

CAT scores. Cognitive Abilities Test (Cognitive Abilities Test, Third Edition (CAT3) http://www.gl-assessment.co.uk) was used in the UK sample. The test as a whole assesses an individual's ability to manipulate and reason with three different types of symbols: words, quantities and spatial patterns. During the complete CAT assessment, test batteries are devoted to each of these ways of reasoning; each battery is further divided into three types of items which test different aspects of that style of reasoning. These results are also used by the secondary school for streaming/tracking students in their maths classes. The UK participants completed these tests either at primary school in year 6 or on a visit to their secondary school prior to commencement of year 7. Data were made available to the study.

Subject achievement. In the UK, National Curriculum assessments for all subjects are taken across the academic year at half term and at the end of term. The assessments are taken during the lesson for the corresponding subject and the teachers set and mark the tests. This study used the end of term results that corresponded with the data collections. The tests were scored 1 to 7, a-c, with ‘7’ and ‘a’ being the highest score. These scores were recoded to provide a continuous scale of ability, 1 = lowest, to 21 = highest. UK students’ maths classes are streamed, on the basis of these termly assessments. Students could move up or down a class each half term depending on their test results.

Similarly to the UK, in Russia national guidelines also govern the school curriculum. At the end of the academic year, students receive a mark based on their work throughout the year; this can be in the form of tests and/or coursework. Teachers provide the mark which is graded 1 to 5 where 5 = highest. A grade of 2 and below, which indicates a fail, is very rarely awarded as tests/coursework will be retaken to avoid failure.

Teacher Measures

Teacher demographics and experience. SES Demographic Survey Form (Cobb, 2004) was used to assess teacher SES and experience. Items include gender and date of birth with further questions regarding marital status, children, cultural background, ethnicity and whether or not English/Russian was their first language. Teaching experience and type of teaching was also incorporated. The 21 items of categorical data were used to code/group teachers.

Job satisfaction. The Job Descriptive Index (JDI: Balzer et al., 1997) assessed job satisfaction across four domains. The 63 item measure divided into 4 scales: job interest (18 items), salary (9 items), supervision (18 items) and colleagues (18 items). The items were scored as follows: ‘No’ = 0, ‘Don’t know’ = 1, ‘Yes’ = 3.

Perceptions of classroom atmosphere*.* Classroom Chaos assessed the teacher’s perception of how calm or otherwise they viewed their classroom. The measure was adapted from Confusion, Hubbub and Order Scale (Matheny et al.,1995) which was included in the TEDS teacher booklet when the children were 10 years of age. The measure is the same as that measuring students’ perception but with the inclusion of the following item omitted from the student measure: ‘*We are usually able to stay on top of things; for example, planning activities, getting them ready’*.

Emotional intelligence. Trait Emotional Intelligence Questionnaire (TEIQuE-SF; Petrides & Furnham, 2004) is a short, 30 item, version of the original measure (Petrides, Pérez, & Furnham, 2003). Participants were presented with statements and asked how much they agreed with them using a 7- point scale, 1 = ‘completely disagree’ to 7 = ‘completely agree’. Items included: ‘*expressing emotions with words is not a problem for me’*.

Self-efficacy*.* Teachers’ Sense of Efficacy Questionnaire (Tschannen-Moran & Woolfolk-Hoy, 2001) has two versions, a long form with 24 items and a short form with 12 items. The 12 item version of the measure was used which assessed three subscales: *Instructional strategies; Classroom management; and Student engagement.* Participants were presented with questions/statements and asked to respond using a 9 – point scale, where 1 = ‘completely disagree’ and 9 = ‘completely agree’. Items included: ‘*to what extent are you able to tailor your lessons to the academic level of your students?’.*

Occupational burnout*.* The Maslach Burnout Inventory Educators Survey (Maslach, Jackson, & Leiter, 1996) comprised 22 items that assessed the frequency of the three aspects of burnout experienced by teachers: *Emotional Exhaustion* (feeling emotionally drained from work), *Depersonalization* (impersonal feelings toward students/co-workers), and reduced feelings of *Personal Accomplishment* (feelings of competence and achievement). Participants were presented with statements and asked to rate whether they ever feel this way about their job, using a 7-point scale, where 1= 'never' and 7 = 'everyday'. Items included: '*I feel that I treat some students as if they were impersonal objects'*.

Procedure

Participant consent was obtained via an opt-out form that was sent home to each student's parent/guardian. Those not wishing their child to participate returned the form to exclude them from the pilot study. Verbal consent was also obtained from participants at the beginning of the data collection, and all participants were given the right to withdraw from the study at any time; confidentiality of all participants' responses was also ensured.

Participants took part as a class exercise during their geography lessons; the sample was split across two lesson periods on two separate days. The first class completed the activities under test conditions as the teacher maintained a high level of behaviour control during the lesson. For the second class with another teacher, the same test conditions were difficult to maintain as the teacher had left the room to provide work for several students who had opted out of the study. The concern was that data quality may have been affected for these participants.

After standardised instructions were read to the class, participants were presented with a range of tasks and self-report questionnaires in pencil and paper format. These were contained across two booklets, one for geography and one for maths. The first task to be presented was a newly developed measure to assess geography knowledge in orientation, physical features of the earth and the solar system. The geography measures were prioritised to ensure that these newly adapted instruments were tested within the 1-hour lesson. The maths measures were presented once the students had worked through the geography section. Participants were asked to think of their geography classrooms since the beginning of term for the first booklet, and asked to think of their maths classrooms during the same period for the second booklet. They worked through these activities until the end of the lesson. During this time, the teachers were presented with the teacher measures to work through and give their comments as opposed to providing their data.

Results

The majority of the measures demonstrated adequate reliability in line with the TEDS data (see Table 2.1). The newly adapted geography measures provided similar reliability to the maths measures. However, some of the measures, such as maths/geography tutoring questions, showed low reliability. This may reflect some meaningful differences in response to these items. For example, some students may have extra tutoring because they struggle with the subject, whereas for others it might be that they excel and their parents might want to capitalize on that. In a larger sample, these items can be explored further, stratifying the sample in terms of grades. Geography homework also demonstrated low reliability compared to maths. This may also reflect differences in response to items, for example ‘*my teacher makes useful comments on my homework’;* and *‘I am given interesting work’.* It may indicate that in some schools, the students regularly check their own homework as a class exercise rather than the teacher taking it away for marking and providing feedback.

For the geography problem verification task, it became apparent following the pilot that the number of items needed to be reduced. Thirty-six of the 73 items were removed, reducing the number to 37. These items were excluded on the basis of high levels of incorrect responses due to difficulty of subject matter in relation to the curriculum; duplication of question type; and poor reproduction of maps which made the item difficult to read. Although the Cronbach’s alpha decreased slightly, the reduced item measure is more suitable for this age group. In addition, with a bigger sample of the main study (reported in Chapters 3 to 5), the validity of this measure was adequate: alpha = .85 averaged across waves and between samples.

Table 2.1. Internal validity of adapted measures for geography and their maths counterparts demonstrated by Cronbach's alpha

Measures	No. of Items	N	Cronbach's Alpha
Geography PVT	73	38	.924
Geography PVT revised version	37	38	.883
Self Perceived Geography Ability	3	38	.366
Self Perceived Geography Enjoyment	3	38	.501
Geography Classroom Environment	12	37	.632
Geography Anxiety	9	28	.815
Geography Usefulness	4	37	.467
Geography Homework	5	37	.176
Geography Environment	3	35	.470
Geography Tutoring	3	10	.000
Geography Classroom Chaos	15	38	.458
Perceptions of academic and socioeconomic status	4	31	.701
Theory of Intelligence	3	33	.615
Self Perceived Maths Ability	3	32	.835
Self Perceived Maths Enjoyment	3	31	.904
Maths Classroom Environment	12	36	.916
Maths Homework	5	38	.689
Maths Usefulness	4	33	.748
Maths Environment	3	32	.612
Maths Tutoring	3	10	.402
Maths Classroom chaos	15	36	.738
Maths Anxiety	9	24	.917

Discussion

The main aim of the pilot study was to test the newly developed geography problem verification task and the newly adapted geography classroom measures. The pilot also aimed to assess the timing of the tasks and activities to ensure that testing could be completed within the students' maths lesson. As the study met these requirements, all the measures were maintained in the main study, with a little adjustment to the wording of some items. The geography problem verification task demonstrated excellent reliability and was reduced in number of items to 37. The measures were translated into Russian and small pilots were conducted following the adaptation.

The testing procedure was also revised for the main study, ensuring that teachers had alternative work set for students who might withdraw from the study or finish the tasks earlier within the lesson time. This was to avoid any disruption to the test conditions for other participants. Additionally, based on the pilot, teachers in the main study were asked to remain in the classroom during testing to maintain behaviour, allowing the researchers to collect data under exam conditions.

Chapter 3

Educational settings and academic outcomes: a cross-cultural investigation

Abstract

Research suggests that differences in educational systems underlie differences in academic outcomes (Woessman, 2009). This study investigates academic outcomes for maths and geography in two samples of 11 to 12 year old students from two countries with different education systems. Results show no significant average differences between the two samples for the majority of maths and geography classroom measures, such as performance, motivation, subject anxiety and perceptions of socioeconomic status. A small significant difference was found only for geography performance. This effect was more likely to stem from curricula differences rather than different education systems given the similarity across samples for all other measures. These findings suggest that the two education systems lead to similar educational outcomes, and that factors that drive individual differences within populations are likely to be similar in the UK and Russia.

Introduction

The main focus of this thesis is an investigation of the between class and teacher differences within two countries, the UK and Russia. However, it is important to address any potential variation in academic outcomes that may result from differences between the two countries' education systems and curricula. It has been suggested that variation in institutional structure and tracking underlie differences in student achievement internationally (Woessman, 2009).

Periodically published reports from the Programme for International Student Assessment (PISA) and the Trends in International Mathematics and Science Study (TIMMS) compare mean performance of school students in maths, science and reading between participating countries globally. East Asian countries are consistently highlighted as the top performers in mathematics and science (PISA, 2015; 2012; 2009; TIMSS, 2015; 2011). Subsequently, educational policy makers in other nations aspire to these ranks and continually reassess their own programmes and curricula to increase their countries' mathematics and science performance. The emphasis is on increasing national academic success in order to improve business/career prospects and in turn increase the countries' gross domestic product (GDP).

The UK and Russia take part in TIMMS and PISA, among up to 72 countries. For TIMMS, only England and Northern Ireland participate from the UK, they are ranked separately. This study focuses on the England results as this is the location for the UK schools in the present thesis. TIMMS assesses two age groups, age 9 to 10 and age 13 to 14 years; PISA assesses one age group, age 15 years. In PISA, both the UK and Russia have consistently ranked in average position for average mathematics performance compared with other participating countries (PISA, 2009; 20012). In the most recent report in 2016, using data collected from 72 participating countries in 2015, both the UK and Russia are 'at OECD average' for maths performance, ranking 27 and 23, respectively out of 72 with a mean score of 492 for the UK and a score of 494 for Russia (PISA, 2016). For science, the UK mean has seen an increase to above the OECD average and ranks 15 out of 72 with a mean score of 509. Russia ranks below average for science at 32 out of 72 with a mean score of 487. There is little difference in rank between Russia and the UK for reading, with 26 and 22 respectively, with Russia just below, and the UK just above the OECD average (mean scores of 495 and 498 respectively).

The TIMMS results are slightly different, both the UK and Russia are in the top ten countries in the 2011 assessment (Mullis, Martin, Foy & Arora, 2012). The most recent report on data collected in 2015 from 57 countries and 7 states/provinces, shows Russia in a higher position than England for maths performance at age 9 to 10 years. Russia's score has increased since 2011 to a mean score of 564, whereas England has a mean score of 546 (Mullis, Martin, Foy & Hooper, 2016). At age 13 to 14 years, Russia has a score of 538 and England has a score of 518, the same score as the US. The top performing East Asian countries' scores range from 593 to 618 for the younger age group and 586 to 621 for the older age group (Mullis et al., 2016). Similar results are shown for science, with Russia performing better than England in the younger cohort, 567 vs. 536, but close in average scores for the older cohort, 544 vs. 537. Overall on these tests, Russia and the UK are similar in outcome. They are also largely similar on the TIMMS (2016) survey of maths confidence at age 9 to 10 years. Students' responses on the survey were calculated to give percentages of students who were 'very confident in mathematics', 'confident' and 'not confident'. An average scale score was calculated from the survey responses so that a score of 10.6 and above meant the student was 'very confident', and a score of 8.5 and below indicated 'not confident'; 'confident' students fell between these marks. Russia's maths confidence score dipped to 9.7, an average scale score that was significantly lower than their score in the 2011 survey. England had an average scale score of 10.1, and was not significantly different to their 2011 result. Both countries are only four points apart and are within the range of being 'confident in mathematics'. At age 13 to 14 years, both countries' scores remain in the 'confident' range with the UK at 10.3 and Russia at 9.8. These scores are towards the lower end as a score of 9.7 and below denotes 'not confident' and for 'very confident' the threshold of 12.1 would need to be reached.

Similarity was also found in previous research that investigated motivation in samples from thirteen countries that included Russia and the UK. In that study, self-perceived ability and enjoyment of mathematics were found to be highly similar across all samples (Kovas et al., 2015).

Overall, the similarities between the UK and Russia shown in these studies are surprising, considering a number of differences between the two education systems. One difference is the age at which formal (primary) education commences. In the UK, children begin primary school at 4 to 5 years. Whereas in Russia, primary school begins at age 7 years.

Another difference is school composition in terms of selection or tracking processes. In Russia, students are taught in mixed ability classrooms for all subjects, within mixed ability schools, throughout their education. In mainstream education there is no selection or streaming apart from certain schools that offer specialized curricula, for example, an advanced maths programme for exceptional maths students. In the UK, the policy is different. For primary education, schools and classes are mixed ability but there may be some setting or grouping within these classes. Whereby students within a classroom are grouped to work alongside other children at a similar level of ability. Often, children are grouped together by ability at tables large enough to accommodate several children. In secondary education there are also schools that are mixed ability but the majority of these schools will select students on ability for their maths and English lessons. There are other schools which select students on ability for all subjects, and the students have to pass rigorous tests at age 10 to 11 years before enrollment at age 11 to 12 years. Some districts implement a test at this age for all students to take before they choose their next school. Those who pass will have the opportunity to apply for highly selective schools in the area with a more advanced curriculum, whereas those who fail can only apply for the mixed ability schools. The test has become divisive, separating those who pass and those who fail. Most districts have opted for a more equitable system and stopped testing students in this way. Instead, students can choose from a selection of mixed ability schools in their area, although there will still be some selection for maths and English classes within the school. With or without rigorous testing at this age, the pressure still remains for parents to select the right school for their child.

In Russia, parents have to make this decision at the beginning of their child's schooling as students usually remain in the same school throughout their education, unless they move (e.g. to another city) or enter a specialized school. Generally, students will attend the school most local to home unless they elect a more specialized programme, for example, learning specific languages.

Throughout their school education, students remain within the same class groups to which they are randomly assigned when starting primary school at age 7 years. During primary education, all subjects (with few exceptions, such as second language and physical education) are taught by the same teacher and this teacher also remains with the same class group for the entire four years. When students transition to secondary education at age 11 to 12 years, the existing class groups are randomly allocated to specific teachers for specific subjects. There will be fewer teachers per subject than number of class groups and so for a subject like geography, one geography teacher will teach several classes.

In the UK, although students in primary education will have the same teacher for all their subjects, the teacher will change on a yearly basis. In secondary education, students will have specific subject teachers. UK students will attend a different, larger school at secondary education. Therefore, unlike Russian students, who remain within the same peer group throughout their schooling, UK students will form new class groups with students from other primary schools and perhaps lose most of their primary peer group. For many of their lessons, students will be in the same new groupings, except for maths and English where students' classes are selected on ability, and so they will likely be with a different group of peers for these lessons.

Another difference between the two countries' education systems is the length of the summer break that students are given. In the UK, students finish the school year towards the end of July and return for the next academic year six weeks later in early September. In Russia, students finish the school year towards the end of May and return for the new academic year in early September following a three month break. Despite the disparity in length of summer break, both Russian and UK students fulfill the same number of days schooling throughout the year, they are just distributed differently across the academic year.

Both Russian and UK students have a large amount of change at transition to secondary education. Russian students will no longer have the same teacher that has taught them for the last four years. UK students will no longer go to the same school site they have been attending for the last four years and they will meet many new peers in the new and much larger secondary school. In Russia, educationalists say that the transition can be a huge shock for Russian students and this may affect their performance and motivation. Similarly in the UK, the change of location, teachers and peers may have a large impact on academic outcomes. It is difficult to disentangle these factors from other aspects of the transition. Instead, any decline in performance may be due to a more intensive curriculum that is implemented at secondary education compared to that of primary school; or other factors, such as maturation processes (e.g. Eccles,1999).

The Current Study

In light of the differences between the two countries' education systems, the current study investigates potential differences and similarities between the countries across one academic year on measures of test performance, classroom environment, motivation, attitude towards specific subjects, and subject anxiety, within two domains, maths and geography. Potential differences are also assessed for perceptions of intelligence and socioeconomic status.

The study uses data collected longitudinally across several assessment points over the course of one academic year in four urban schools, two in the UK and two in Russia. All schools are mixed ability, although in the UK, students are streamed by ability for their maths classes. In Russia, the students attend a school where they have the opportunity to learn two second languages. The study addresses the following research questions: 1) Are there differences between the two countries in academic outcomes? 2) Do potential differences persist across the academic year? 3) Are the patterns of results similar for maths and geography?

Methods

Participants

Participants were 520 10 to 12 year old students, from four urban mixed ability schools; two in London, UK and two in St. Petersburg, Russia (see Table 3.1 for sample characteristics). Although the UK schools were mixed ability, students were streamed by ability for their maths classes. The Russian students were not streamed for ability. However, they attended schools with specialized linguistic programmes that provided the students with the opportunity to learn up to two languages: English; English and Spanish; and English and Chinese. In one school, there were eight classes of students who learned English and/or Spanish. In the other school, there were three classes of students who learned English and Chinese. Previous research with another cohort of students from the same school shows no differences between the students following different language programmes on cognitive tests suggesting similarity in ability across the linguistic groups following one year of learning different second languages (Rodic et al., 2015).

All students were in the first year of their secondary education, with specific subject teachers for the first time. Students with special educational needs were excluded from these analyses.

Table 3.1. Sample characteristics for the UK and Russian students at each assessment wave: gender, mean age in months and standard deviation (SD), and N

		Time 1	Time 2	Time 3
Russia	Male	102	99	98
	Female	127	125	129
	Total n	229	224	227
	Mean age (months)	139.29	142.60	146.77
	SD	4.27	4.19	4.04
	Minimum	127	131	135
	Maximum	148	153	156
UK	Male	152	151	163
	Female	131	132	130
	Total n	283	283	293
	Mean age (months)	140.98	144.53	149.99
	SD	3.81	3.69	3.75
	Minimum	135	139	143
	Maximum	158	156	163
N	Total	512	507	520

Measures

A detailed description of the measures used in this study is provided in the methods section in Chapter 2, pages 61 to 70.

Procedure

The procedure was standardised across both countries so that all data collections followed the same format.

Participant consent was obtained via an opt-out form that was sent home to each student’s parent/guardian. Those not wishing their child to participate returned the form to exclude them from the study. Verbal consent was obtained from participants at the beginning of each data collection, and all participants were given the right to withdraw from the study at any time. Confidentiality of all participants’ responses was also ensured.

Participants took part as a class exercise during their mathematics lessons under test conditions. In Russia data were collected at three assessment points: the first - at the beginning of the spring term; the second - in April/May at the end of the school year; and the third in September when students returned from their summer break (see Figure 3.1). At each assessment, up to two classes were tested per day so data collection took place over the course of two weeks. In the UK, data were collected at five assessment points: the first was at the beginning of the academic year; the second - at the end of the autumn term (December); the third was in March/April, at the end of the spring term; the fourth was in July, the end of the summer term; the final collection was in September, at the start of the new academic year following their summer break (See Figure 3.2). The data collection in the UK also took place over the course of two weeks, data were collected from half of the classes in a year group in one sitting at each school.

After standardised instructions were read to the class, participants were presented with a range of tasks and self-report questionnaires in pencil and paper format. The first task to be presented was the Maths Problem Verification task (MPVT), which is a timed test. Eight minutes were allowed for completion of the task, following this, papers were collected to prevent participants returning to unfinished items. The participants were given the remainder of the lesson to complete the rest of the activities.

The non-cognitive measures were grouped and presented separately for each subject. Participants were asked to think about their maths classrooms since the beginning of term for the first eight measures, and asked to think about their geography classrooms for the last eight measures.

While the students participated, data were also collected from their teachers for use in other analyses. These data were collected at the first assessment in both countries and at the fourth assessment in the UK.

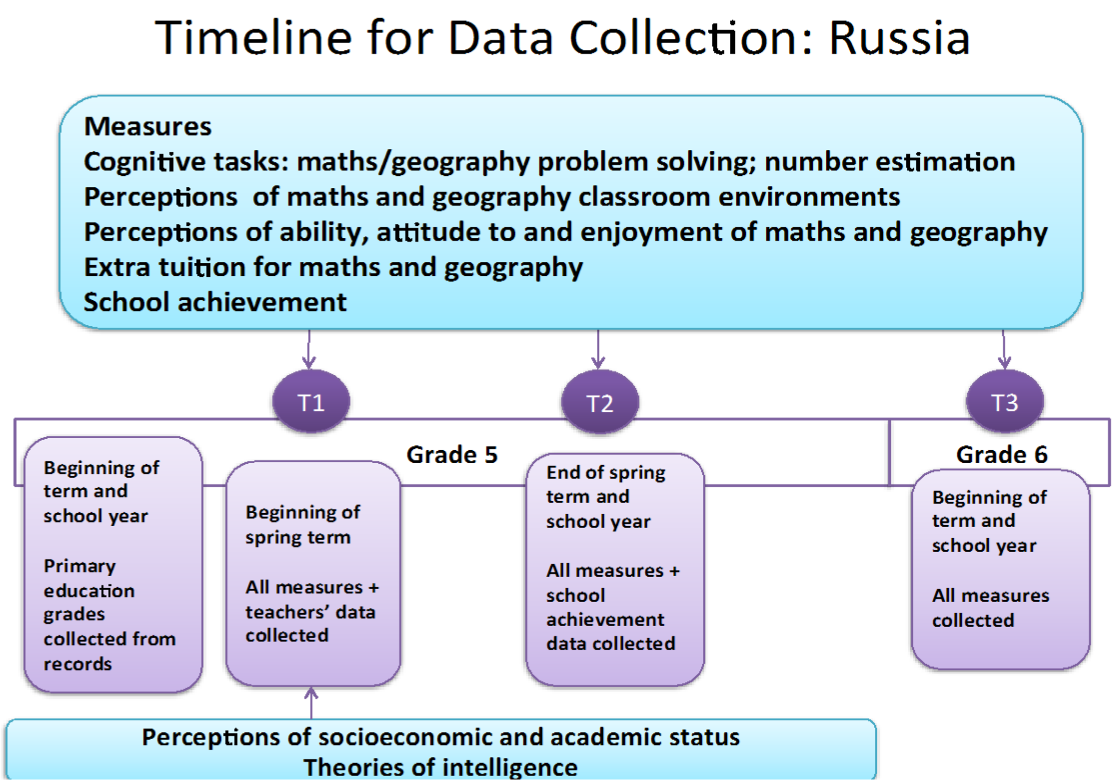

Figure 3.1. Timeline of data collection for the Russian sample (T1: January; T2: April/May; T3: September)

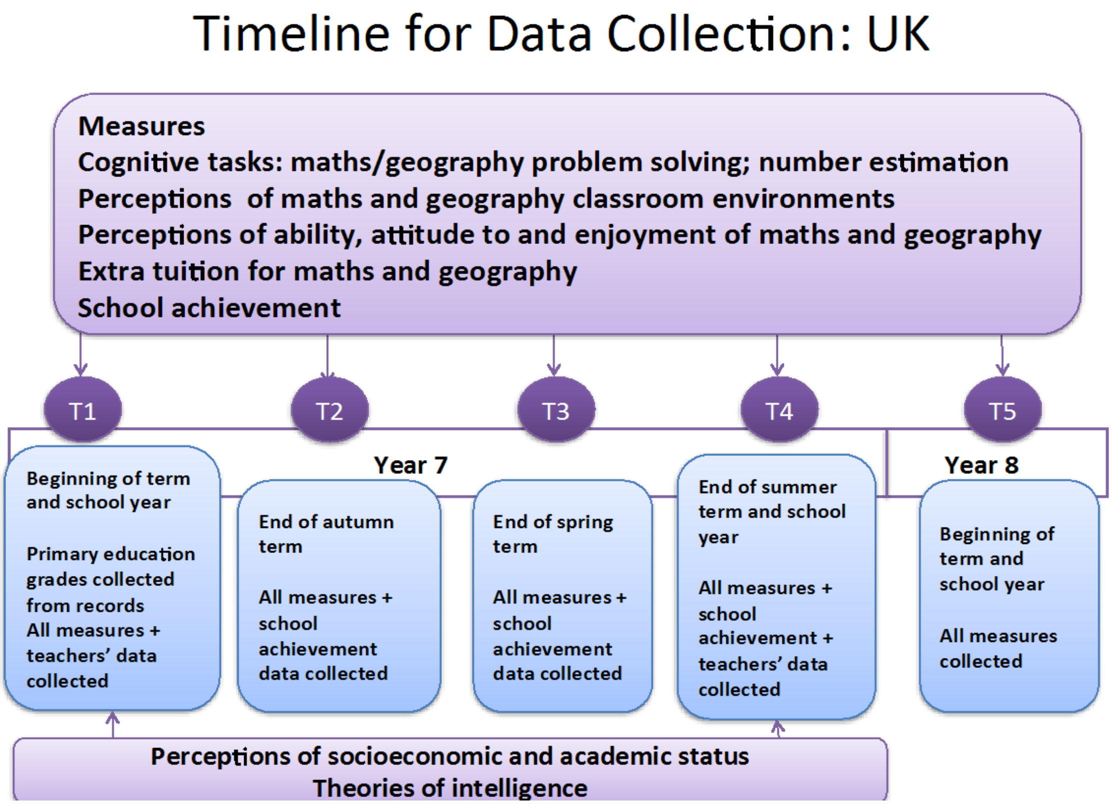

Figure 3.2. Timeline of data collection for the UK sample (T1: September; T2 December; T3 March/April; T4: July; T5 September)

Analyses

Analyses were conducted using data collected from the UK schools at time 2, 3 and 5, corresponding with the data collections in Russia at time 1, 2, and 3. Prior to analyses, variables were tested for normality to ensure their suitability for use with parametric tests. Transformed number line task, geography performance, and homework behavior for both subjects, were used in these analyses as skewness occurred at different waves in one or both samples. Variables were also corrected for age and outliers (±3SD) were removed.

Prior to the main analyses, bivariate correlations were conducted on all variables collected at each assessment to assess their stability across the academic year.

Mixed analyses of variance (ANOVA) were conducted separately for maths and geography measures by country; and time (1, 2 and 3, as described above) by country. They were conducted to assess potential differences in means and variance for maths and geography performance, classroom environment, motivation, attitude towards subject, subject anxiety, and perceptions of intelligence and socioeconomic status. A Bonferroni multiple testing correction was set of $p \leq .001$ where $p = .05$ divided by the number of measures (k=90) across maths and geography and across the three measurement points. This translates as: maths classroom measures = 14 x 3; geography classroom measures = 13 x 3; maths achievement =1 x 2 (time 1 and time 2 only); geography achievement = 1 x 1 (time 2 only); perceptions of intelligence and socioeconomic status = 6 x 1 (time 1 only).

Results

Descriptive statistics for all assessed variables for each sample are presented in Appendix 1 (Table 1.6 for maths, Table 1.7 for geography and Table 1. 8 for perceptions of intelligence and socioeconomic status). The stability of all assessed variables across the three assessment points for the whole sample combined are presented in Tables 3.2.1 to 3.2.3 for maths and Tables 3.3.1 to 3.3.3 for geography

Table 3.2.1. Stability of maths classroom measures across time 1, time 2 and time 3 for the UK and Russian sample combined (N)

	Maths performance	Number line	Maths self-perceived ability	Maths enjoyment
Time 1	1 (519)	1 (514)	1 (504)	1 (494)
Time 2	.670** (471)	.423** (462)	.678** (443)	.589** (438)
Time 3	.672** (465)	.409** (458)	.625** (443)	.578** (429)

Stability was estimated with bivariate correlations **. Correlation is significant at the 0.01 level (2-tailed) *. Correlation is significant at the 0.05 level (2-tailed).

Table 3.2.2. Stability of maths classroom measures across time 1, time 2 and time 3 for the UK and Russian sample combined (N)

	Maths classroom environ-ment	Maths classroom student-teacher relations	Maths classroom peer com-petition	Maths classroom chaos	Maths homework behaviour
Time 1	1 (513)	1 (513)	1 (513)	1 (517)	1 (516)
Time 2	.467** (462)	.462** (462)	.433** (459)	.525** (468)	.652** (465)
Time 3	.384** (453)	.373** (455)	.345** (458)	.457** (464)	.508** (461)

Stability was estimated with bivariate correlations **. Correlation is significant at the 0.01 level (2-tailed) *. Correlation is significant at the 0.05 level (2-tailed).

Table 3.2.3. Stability of maths classroom measures across time 1, time 2 and time 3 for the UK and Russian sample combined (N)

	Maths homework feedback	Maths homework total scale	Maths environ-ment	Maths usefulness	Maths anxiety
Time 1	1 (515)	1 (515)	1 (507)	1 (507)	1 (508)
Time 2	.507** (460)	.576** (462)	.409** (452)	.449** (448)	.533** (448)
Time 3	.358** (456)	.405** (456)	.297** (446)	.450** (448)	.521** (450)

Stability was estimated with bivariate correlations **. Correlation is significant at the 0.01 level (2-tailed) *. Correlation is significant at the 0.05 level (2-tailed).

Table 3.3.1. Stability of geography classroom measures across time 1, time 2 and time 3 for the UK and Russian sample combined (N)

	Geography performance	Geography self-perceived ability	Geography enjoyment	Geography classroom environment
Time 1	1 (515)	1 (477)	1 (483)	1 (476)
Time 2	.564** (466)	.559** (413)	.497** (428)	.436** (421)
Time 3	.623** (462)	.549** (418)	.544** (419)	.318** (418)

Stability was estimated with bivariate correlations **. Correlation is significant at the 0.01 level (2-tailed) *. Correlation is significant at the 0.05 level (2-tailed).

Table 3.3.2. Stability of geography classroom measures across time 1, time 2 and time 3 for the UK and Russian sample combined (N)

	Geography classroom student-teacher relations	Geography classroom peer com-petititon	Geography classroom chaos	Geography homework behaviour	Geography homework feedback
Time 1	1 (476)	1 (476)	1 (478)	1 (477)	1 (476)
Time 2	.411** (422)	.357** (422)	.515** (418)	.588** (417)	.529** (416)
Time 3	.369** (420)	.280** (420)	.394** (419)	.444** (412)	.350** (409)

Stability was estimated with bivariate correlations **. Correlation is significant at the 0.01 level (2-tailed) *. Correlation is significant at the 0.05 level (2-tailed).

Table 3.3.3. Stability of geography classroom measures across time 1, time 2 and time 3 for the UK and Russian sample combined (N)

	Geography homework total scale	Geography environment	Geography usefulness	Geography anxiety
Time 1	1 (476)	1 (459)	1 (465)	1 (473)
Time 2	.570** (416)	.302** (392)	.338** (406)	.530** (414)
Time 3	.387** (409)	.180** (394)	.226** (399)	.545** (415)

Stability was estimated with bivariate correlations **. Correlation is significant at the 0.01 level (2-tailed) *. Correlation is significant at the 0.05 level (2-tailed).

Cross-Country Comparisons at Time 1, Time 2 and Time 3

Maths classroom measures. Figures 3.3.1 to 3.3.14 present the trajectory of means with standard errors for all assessed maths variables across the three assessments by country. ANOVA results for maths classroom measures by country and by country and time are presented in Table 3.4.The results show for all measures, no significant main effect of country, no significant main effect of time, and no significant interaction of country by time following a multiple testing correction of $p \leq .001$($p = .05/90$).

Results from Levene's tests showed that equal variance could be assumed for the majority of analyses apart from the number line task at time 2 ($p < .001$), and at time 3 ($p = .001$), student-teacher relations at time 3 ($p = .024$) and maths usefulness at time 1 ($p = .027$) (see Appendix 3, Tables 3.1 and 3.2). Mauchly's test results also indicated that sphericity could be assumed for almost all analyses apart from maths homework behaviour, χ^2 (2) = 17.500, $p < .001$ (see Appendix 3, Table 3.3), Greenhouse-Geisser results were reported for these analyse

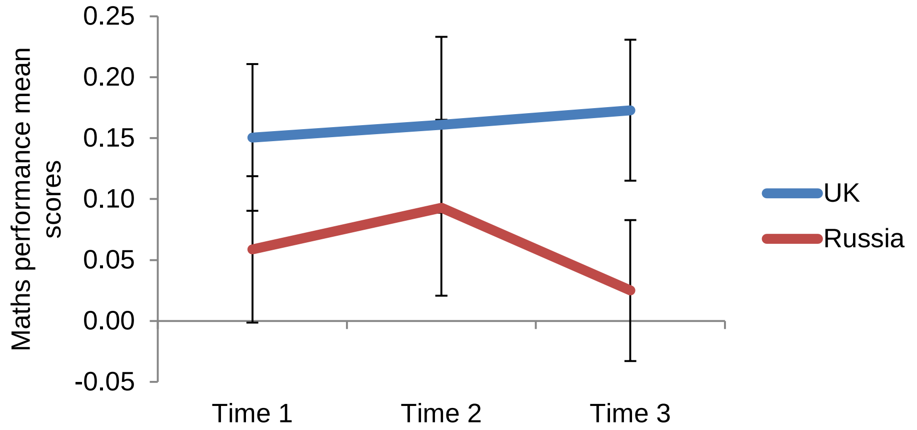

Figure 3.3.1. Means and standard errors for maths performance at time 1, time 2 and time 3 for UK and Russia.

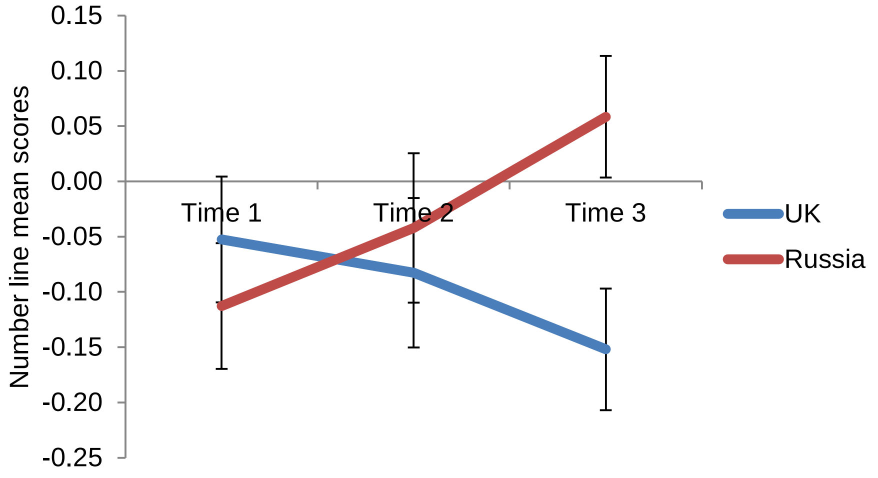

Figure 3.3.2. Means and standard errors for Number line at time 1, time 2 and time 3 for UK and Russia. Note: Unequal variances were shown at time 2 and time 3. The smallest variance at time 2 was shown for the UK (0.66) and the largest for Russia (0.85). At time 3, the smallest variance was shown for the UK (0.67) and the largest for Russia (0.98).

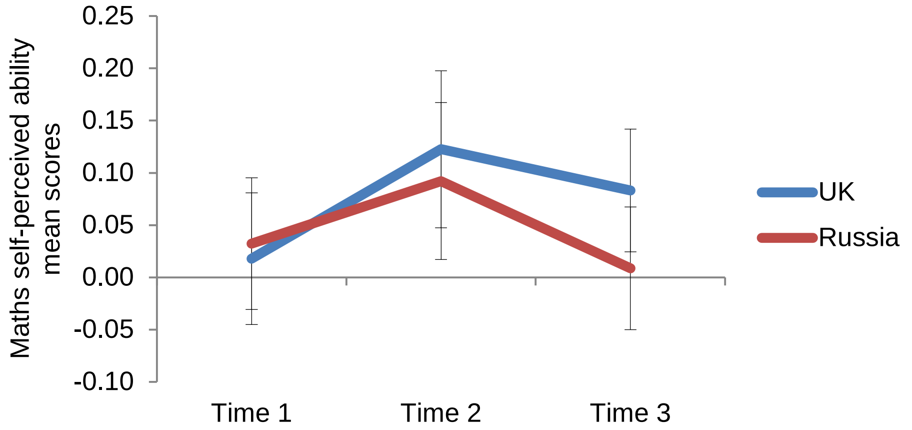

Figure 3.3.3. Means and standard errors for maths self-perceived ability at time 1, time 2 and time 3 for UK and Russia.

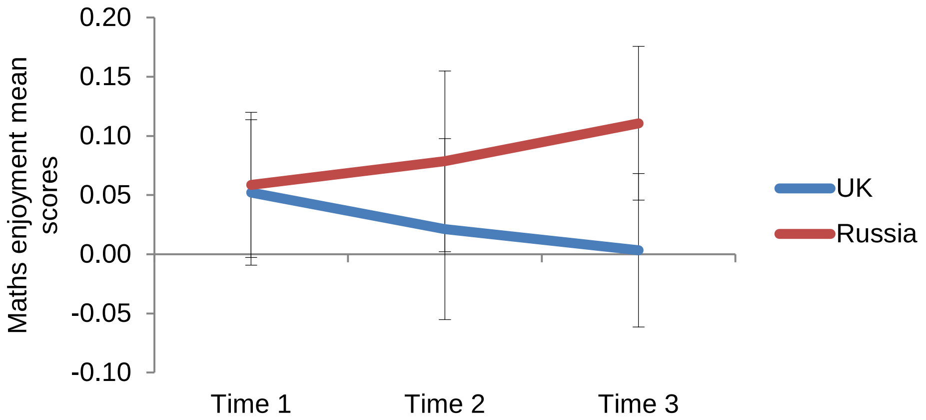

Figure 3.3.4. Means and standard errors for maths enjoyment at time 1, time 2 and time 3 for UK and Russia.

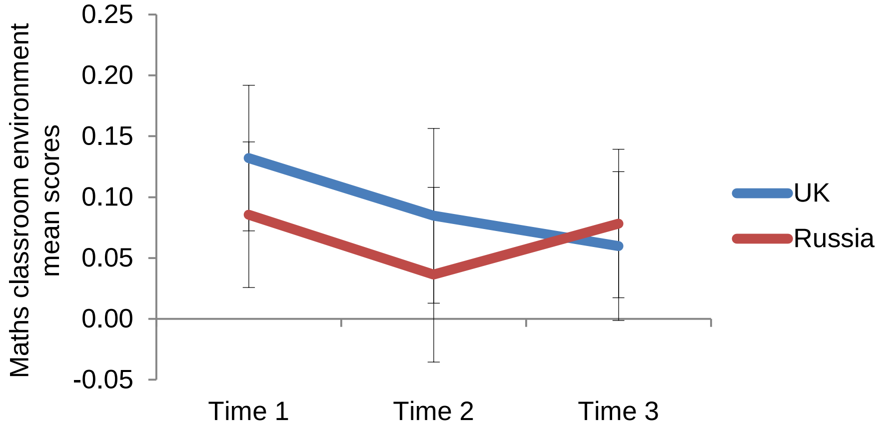

Figure 3.3.5. Means and standard errors for maths classroom environment at time 1, time 2 and time 3 for UK and Russia.

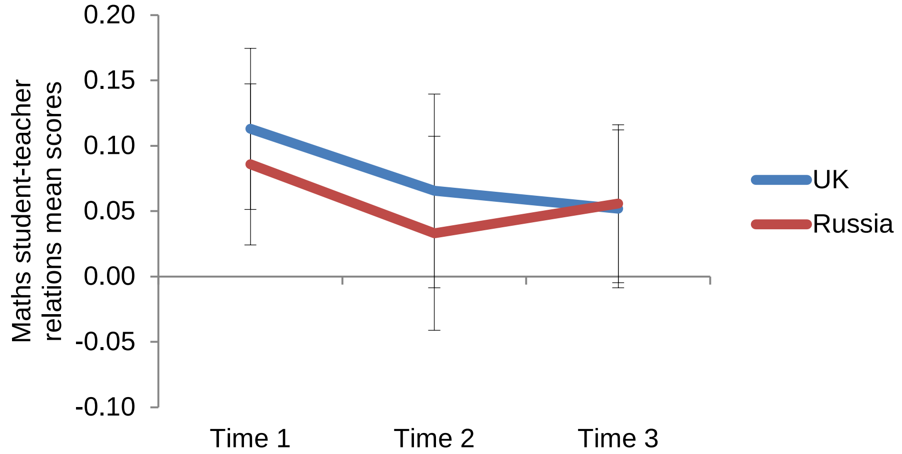

Figure 3.3.6. Means and standard errors for maths student-teacher relations at time 1, time 2 and time 3 for UK and Russia. Note: Unequal variances were shown at time 3. The smallest variance was shown for Russia (0.75) and the largest for the UK (1.02).

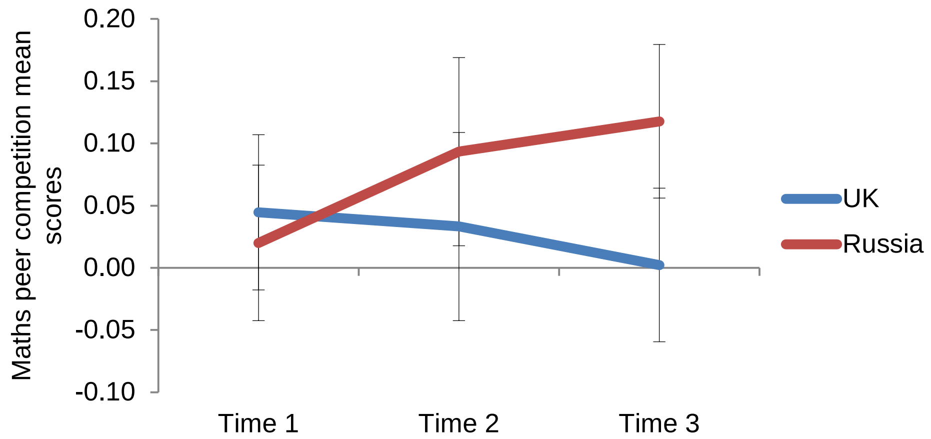

Figure 3.3.7. Means and standard errors for maths peer competition at time 1, time 2 and time 3 for UK and Russia.

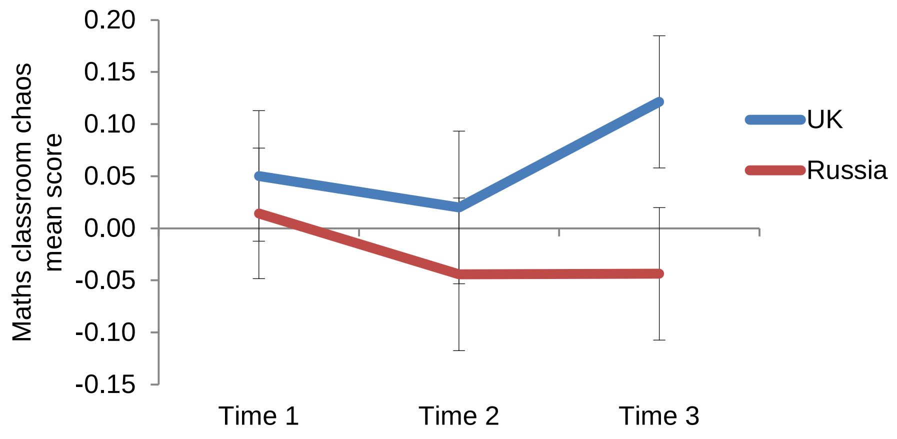

Figure 3.3.8. Means and standard errors for maths classroom chaos at time 1, time 2 and time 3 for UK and Russia. Note: a high score indicates low chaos.

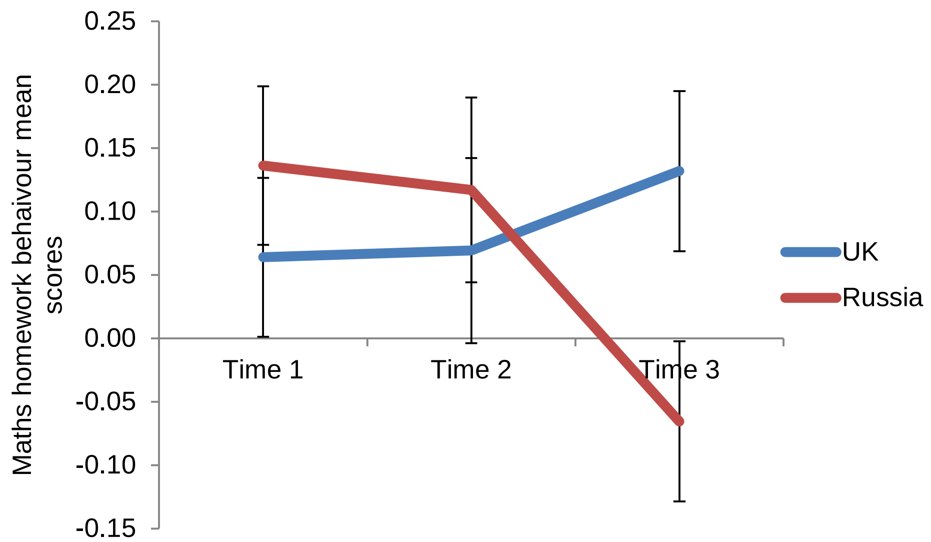

Figure 3.3.9. Means and standard errors for maths homework behaviour at time 1, time 2 and time 3 for UK and Russia. Note: the assumption of sphericity was violated for these analyses (see Appendix 3, Table 3.3).

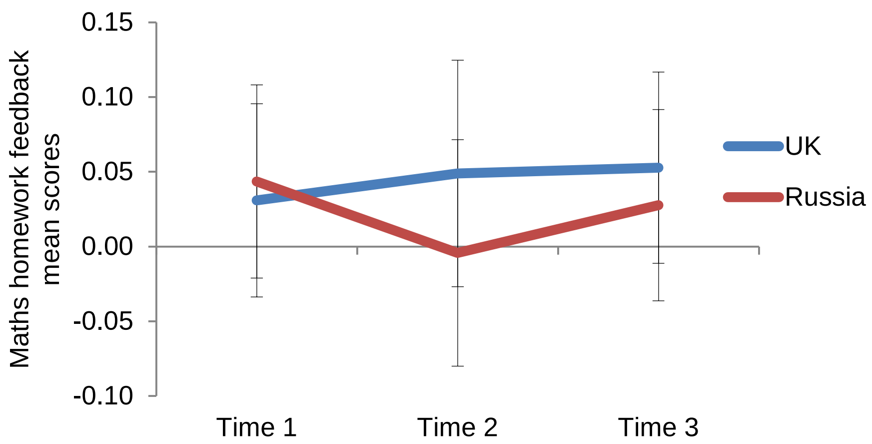

Figure 3.3.10. Means and standard errors for maths homework feedback at time 1, time 2 and time 3 for UK and Russia.

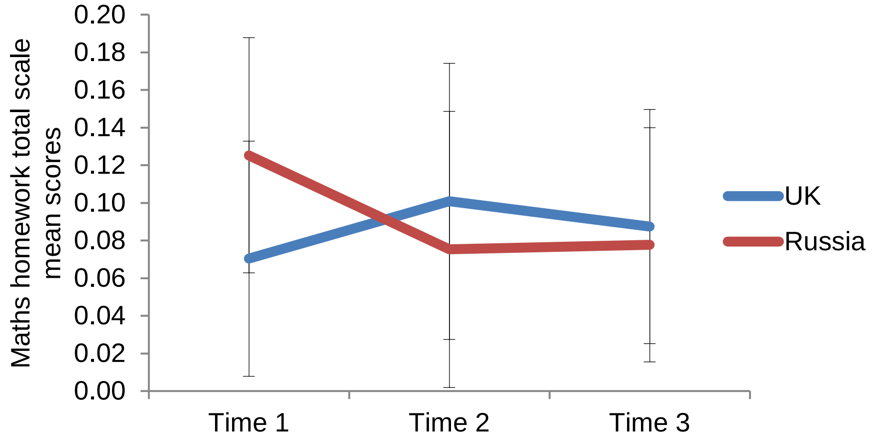

Figure 3.3.11. Means and standard errors for maths homework total scale at time 1, time 2 and time 3 for UK and Russia.

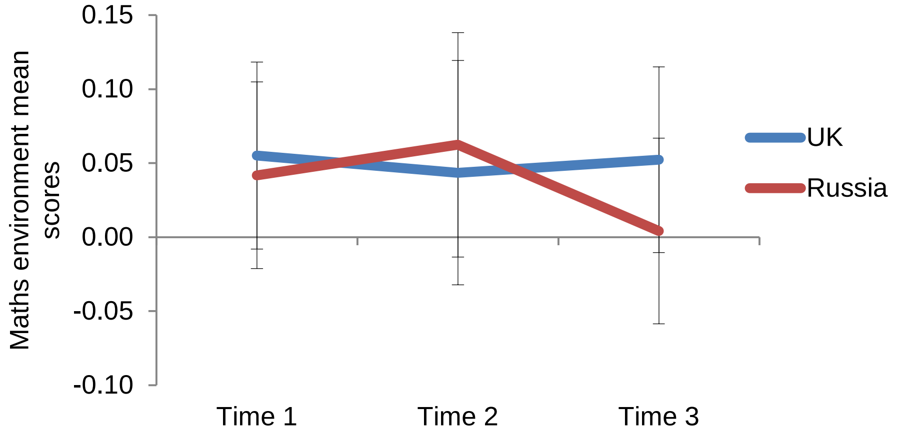

Figure 3.3.12. Means and standard errors for maths environment at time 1, time 2 and time 3 for UK and Russia.

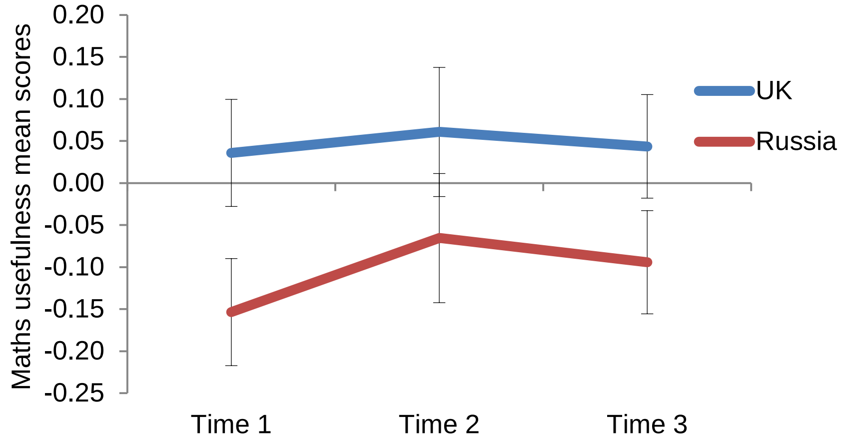

Figure 3.3.13. Means and standard errors for maths usefulness at time 1, time 2 and time 3 for UK and Russia. Note: Unequal variances were shown at time 1. The smallest variance was shown for Russia (0.77) and the largest for the UK (1.08).

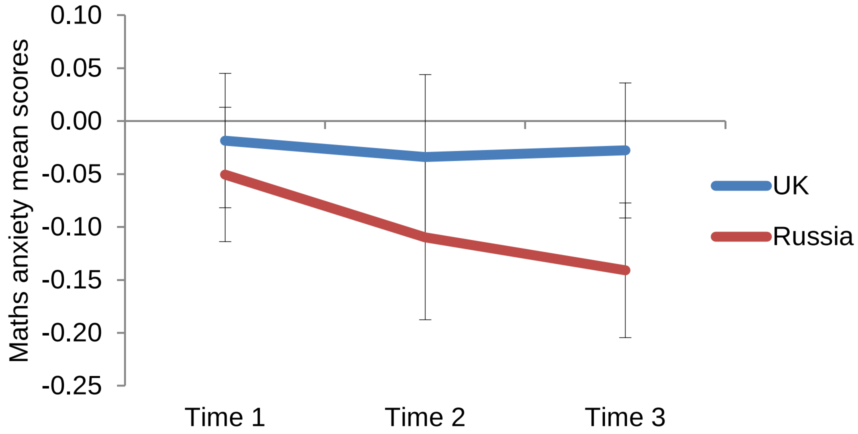

Figure 3.3.14. Means and standard errors for maths anxiety at time 1, time 2 and time 3 for UK and Russia.

Table 3.4. ANOVA results for maths classroom measures by country and time, across time 1, time 2 and time 3

Construct	**Effects**	**df**	***F***	***P***	η_p^2
Maths performance	time	2,846	.330	.719	.001
	time * Country	2,846	.628	.534	.001
	Country	1,423	1.551	.214	.004
Number line	time	2,826	.290	.748	.001
	time * Country	2,826	4.213	.015	.010
	Country	1,413	.855	.356	.002
Maths self-perceived ability	time	2,784	2.214	.110	.006
	time * Country	2,784	.599	.549	.002
	Country	1,392	.134	.715	.000
Maths enjoyment	time	2,762	.013	.987	.000
	time * Country	2,762	.597	.550	.002
	Country	1,381	.440	.508	.001
Maths classroom environment	time	2,824	.533	.587	.001
	time * Country	2,824	.291	.748	.001
	Country	1,412	4.646	.032	.011
Maths classroom student-teacher relations	time	2,828	.588	.555	.001
	time * Country	2,828	.073	.929	.000
	Country	1,414	.062	.804	.000
Maths classroom peer competition	time	2,824	.212	.809	.001
	time * Country	2,824	.908	.404	.002
	Country	1,412	.444	.505	.001
Maths classroom chaos	time	2,842	.682	.503	.002
	time * Country	2,842	1.025	.358	.002
	Country	1,421	1.228	.268	.003

Significant results in bold following a Bonferroni multiple testing correction of $p \leq .001$ ($p = .05$ divided by number of measures (k=90) across maths and geography, time 1, 2 & 3). *Assumption of sphericity violated, Greenhouse-Geisser results reported.

Table 3.4. continued. ANOVA results for maths classroom measures by country and time, across time 1, time 2 and time 3

Construct	Effects	df	F	P	ηp2
Maths homework behaviour*	time	2,836	1.395	.248	.003
	time * Country	2,836	5.730	.004	.014
	Country	1,418	.097	.756	.000
Maths homework feedback	time	2,818	.066	.933	.000
	time * Country	2,818	.195	.818	.000
	Country	1,409	.077	.781	.000
Maths homework total scale	time	2,824	.053	.945	.000
	time * Country	2,824	.399	.666	.001
	Country	1,412	.007	.934	.000
Maths environment	time	2,802	.110	.896	.000
	time * Country	2,802	.178	.837	.000
	Country	1,401	.037	.848	.000
Maths usefulness	time	2,798	.618	.539	.002
	time * Country	2,798	.216	.806	.001
	Country	1,399	3.749	.054	.009
Maths anxiety	time	2,798	.599	.547	.001
	time * Country	2,798	.368	.689	.001
	Country	1,399	.778	.378	.002

Significant results in bold following a Bonferroni multiple testing correction of $p \leq .001$ ($p = .05$ divided by number of measures (k=90) across maths and geography, time 1, 2 & 3). *Assumption of sphericity violated, Greenhouse-Geisser results reported.

Geography classroom measures. Figures 3.4.1 to 3.4.13 present the trajectory of means with standard errors for assessed variables across the three assessment waves by country. ANOVA results for geography classroom measures by country, and by country and time, are presented in Table 3.5. The results show for the majority of measures, no significant main effect of country, no significant main effect of time, and no significant interaction of country by time following a multiple testing correction of $p \leq .001$ ($p = .05/90$). Results from Levene's tests showed that equal variance could be assumed for all analyses apart from geography performance at time 1 and time 3 ($p \leq .014$ and $p < .001$, respectively) (see Appendix 3, Tables 3.4 and 3.5). Mauchly's test results showed that sphericity could be assumed for all analyses apart from geography homework behaviour and geography homework total scale (see Appendix 3,

Table 3.6).

The only measure showing a significant difference was geography performance, which demonstrated a small significant main effect of country, $F(1, 419) = 22.877$, $p < .001$, $\eta_p^2 = .052$. Figure 3.4.1 below shows that on average across the three waves, students in the UK sample performed significantly better than students in the Russian sample (see Table 3.5). However, Levene's test revealed unequal variances for these analyses, with a larger amount of variance shown in the UK sample compared to the Russian sample at time 1 (0.92 vs 0.71) and at time 3 (1.04 vs. 0.64).

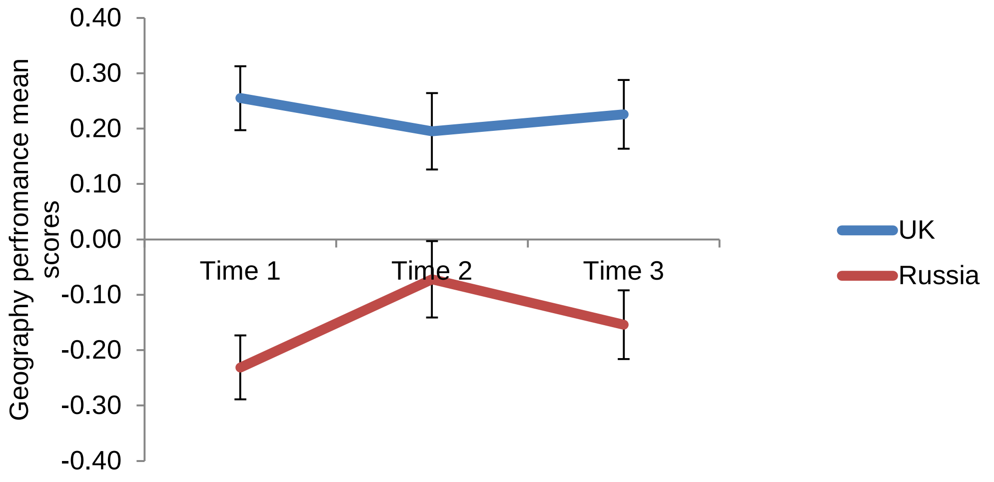

Figure 3.4.1. Means and standard errors for geography performance at time 1, time 2 and time 3 for UK and Russia. Note: Unequal variances were shown at time 1 and time 3 (see above).

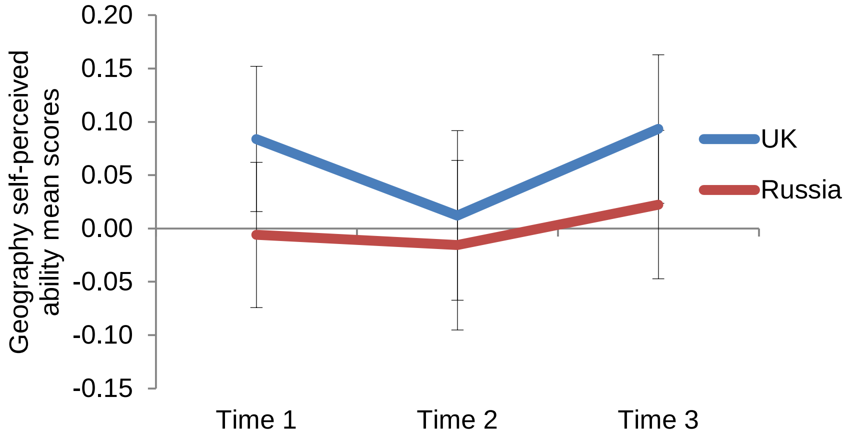

Figure 3.4.2. Means and standard errors for geography self-perceived ability at time 1, time 2 and time 3 for UK and Russia.

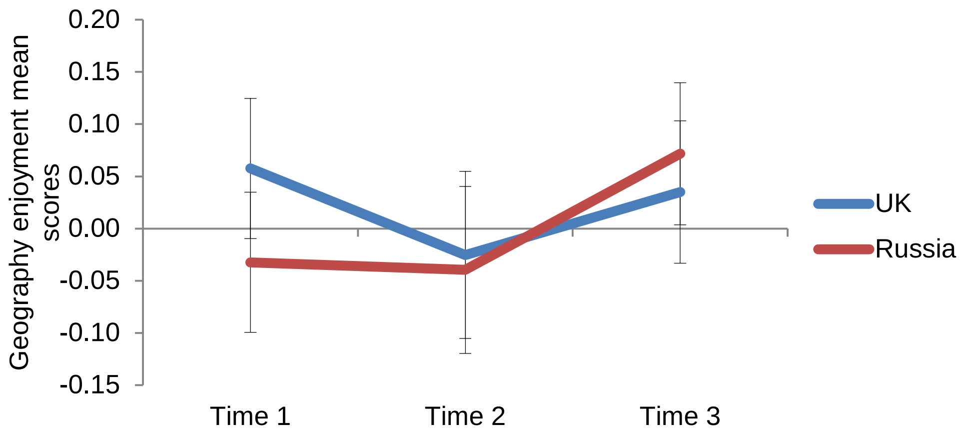

Figure 3.4.3. Means and standard errors for geography enjoyment at time 1, time 2 and time 3 for UK and Russia.

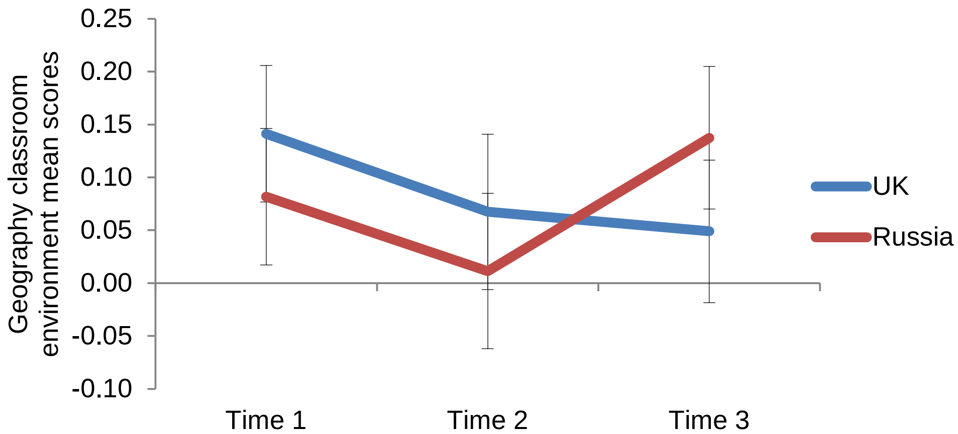

Figure 3.4.4. Means and standard errors for geography classroom environment at time 1, time 2 and time 3 for UK and Russia.

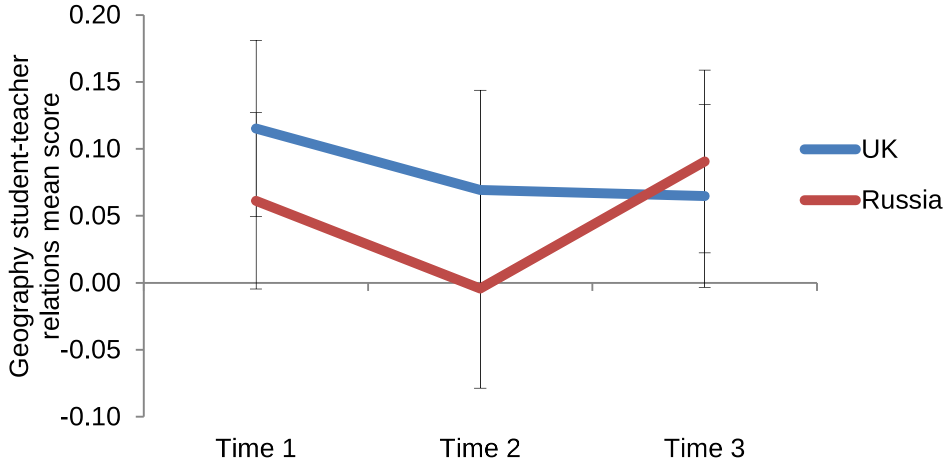

Figure 3.4.5. Means and standard errors for geography student-teacher relations at time 1, time 2 and time 3 for UK and Russia.

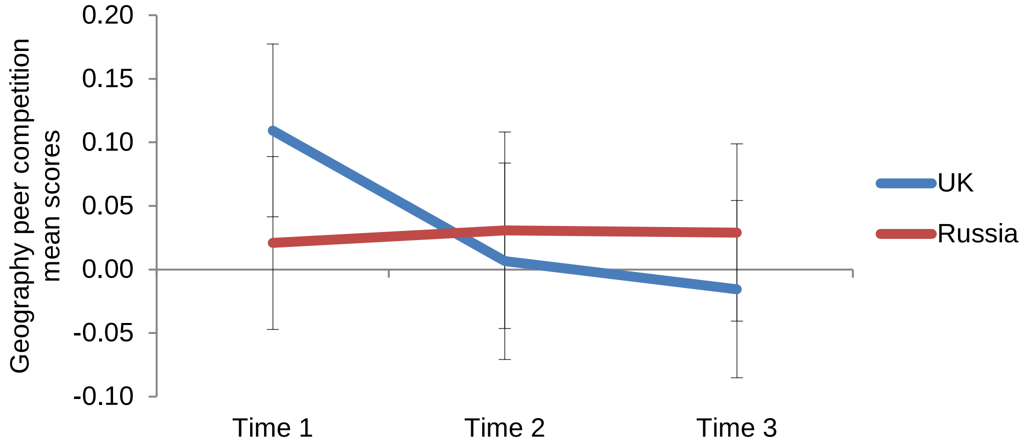

Figure 3.4.6. Means and standard errors for geography peer competition at time 1, time 2 and time 3 for UK and Russia.

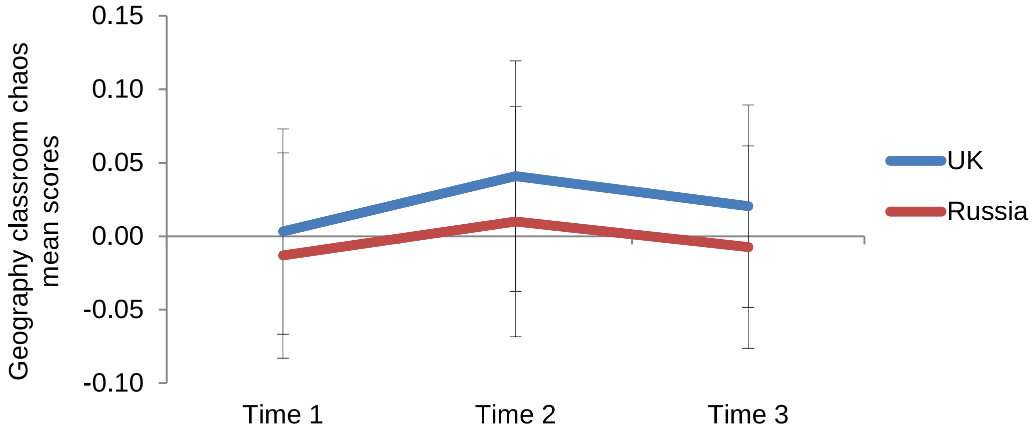

Figure 3.4.7. Means and standard errors for geography classroom chaos at time 1, time 2 and time 3 for UK and Russia. Note: a high score indicates low chaos.

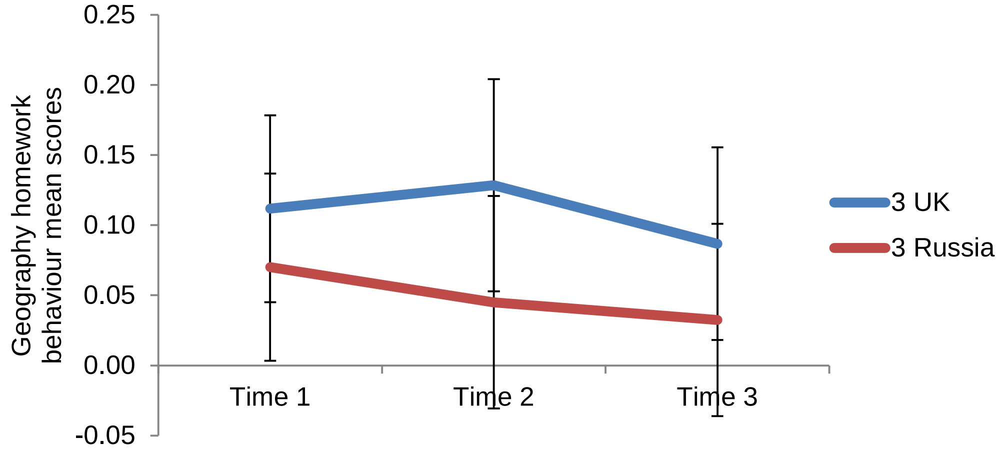

Figure 3.4.8. Means and standard errors for geography homework behaviour at time 1, time 2 and time 3 for UK and Russia. Note: the assumption of sphericity was violated for these analyses (see Appendix 3, Table 3.6).

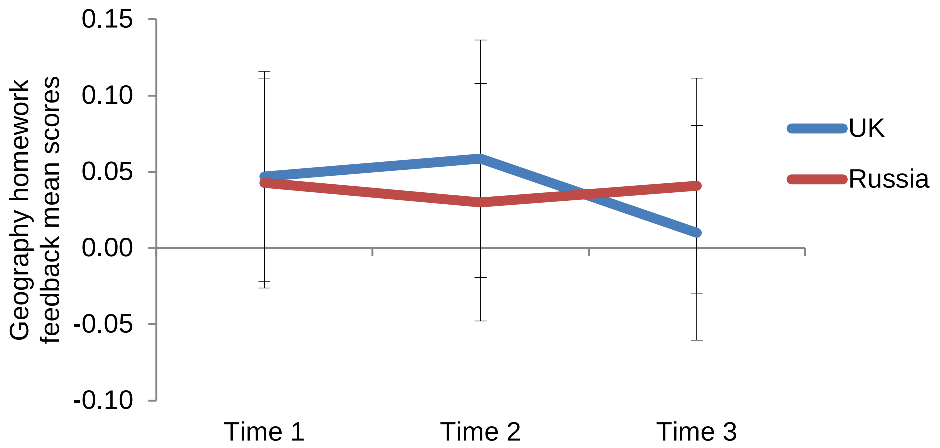

Figure 3.4.9. Means and standard errors for geography homework feedback at time 1, time 2 and time 3 for UK and Russia.

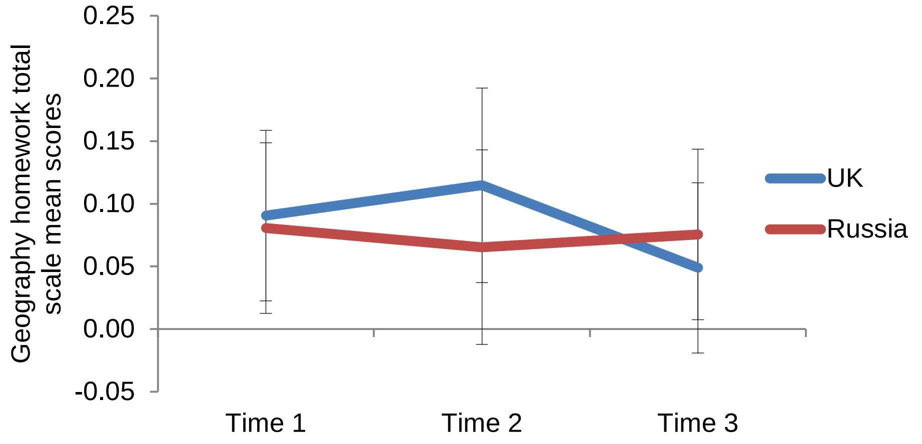

Figure 3.4.10. Means and standard errors for geography homework total scale at time 1, time 2 and time 3 for UK and Russia. Note: the assumption of sphericity was violated for these analyses (see Appendix 3, Table 3.6).

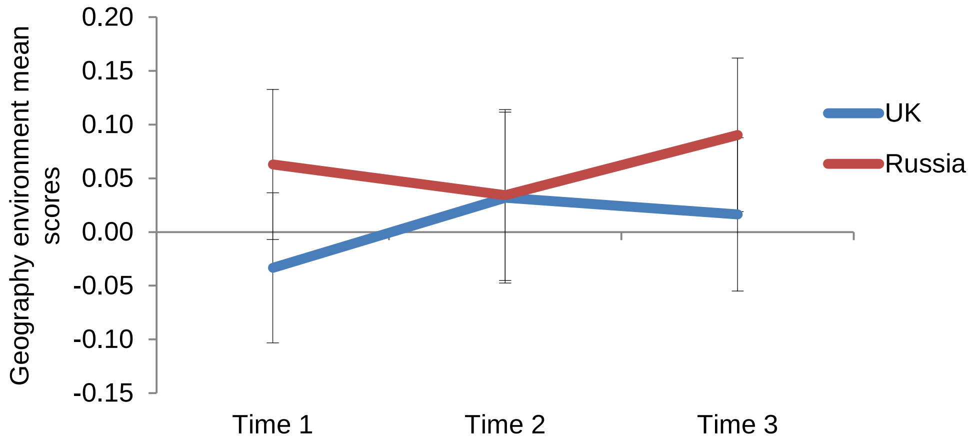

Figure 3.4.11. Means and standard errors for geography environment at time 1, time 2 and time 3 for UK and Russia.

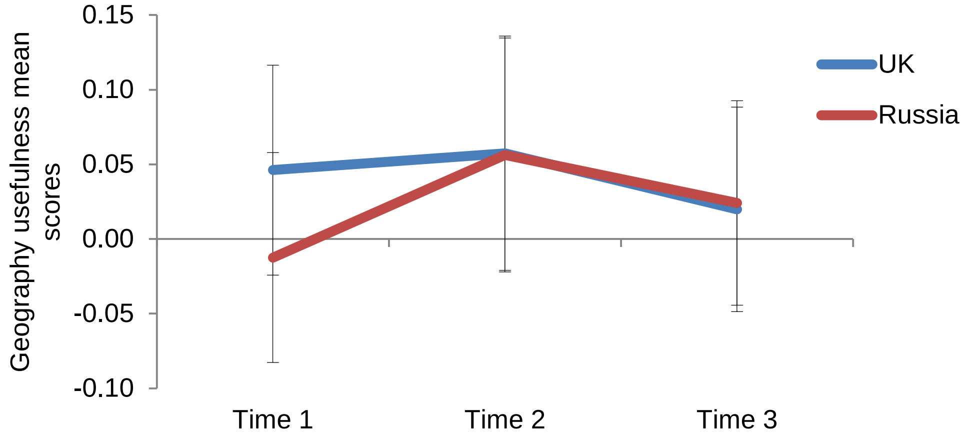

Figure 3.4.12. Means and standard errors for geography usefulness at time 1, time 2 and time 3 for UK and Russia.

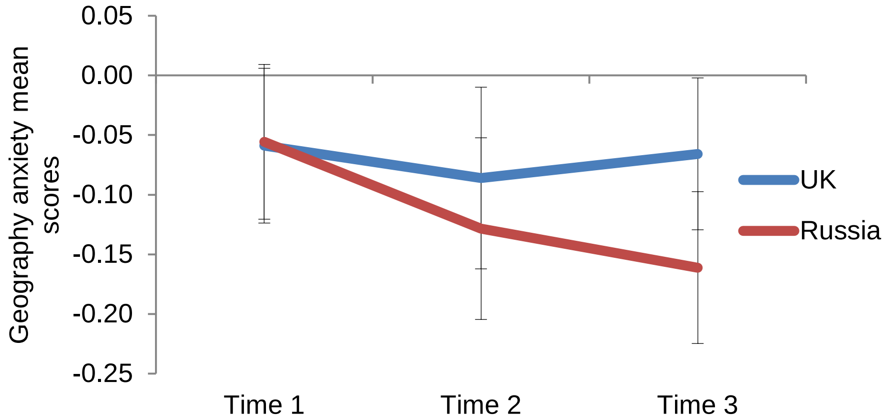

Figure 3.4.13. Means and standard errors for geography anxiety at time 1, time 2 and time 3 for UK and Russia.

Table 3.5. ANOVA results for geography classroom measures by country and time, across time 1, time 2 and time 3

Construct	Effects	df	*F*	*p*	η_p^2
Geography performance	time	2,838	.685	.504	.002
	time * Country	2,838	3.336	.036	.008
	Country	1,419	22.877	**.000**	**.052**
Geography self-perceived ability	time	2,734	.735	.480	.002
	time * Country	2,734	.202	.818	.001
	Country	1,367	.518	.472	.001
Geography enjoyment	time	2,752	1.465	.232	.004
	time * Country	2,752	.806	.447	.002
	Country	1,376	.064	.800	.000
Geography classroom environment	time	2,744	.852	.425	.002
	time * Country	2,744	1.082	.338	.003
	Country	1,372	.014	.905	.000
Geography classroom student-teacher relations	time	2,750	.525	.592	.001
	time * Country	2,750	.414	.661	.001
	Country	1,375	.192	.662	.001
Geography classroom peer competition	time	2,748	.512	.599	.001
	time * Country	2,748	.693	.500	.002
	Country	1,374	.007	.933	.000
Geography classroom chaos	time	2,738	.159	.849	.000
	time * Country	2,738	.010	.989	.000
	Country	1,369	.091	.763	.000
Geography homework behaviour	time	2,728	.230	.790	.001
	time * Country	2,728	.091	.909	.000
	Country	1,364	.499	.481	.001
Geography homework feedback	time	2,722	.076	.927	.000
	time * Country	2,722	.141	.869	.000
	Country	1,361	.000	.994	.000
Geography homework total scale	time	2,722	.157	.850	.000
	time * Country	2,722	.253	.771	.001
	Country	1,361	.017	.896	.000
Geography environment	time	2,686	.178	.837	.001
	time * Country	2,686	.287	.751	.001
	Country	1,343	.539	.463	.002
Geography usefulness	time	2,700	.243	.784	.001
	time * Country	2,700	.156	.856	.000
	Country	1,350	.059	.808	.000
Geography anxiety	time	2,730	.871	.419	.002
	time * Country	2,730	.559	.572	.002
	Country	1,365	.292	.589	.001

Significant results in bold following a Bonferroni multiple testing correction of $p \leq .001$ ($p = .05$ divided by number of measures (k=90) across maths and geography, time 1, 2 & 3). *Geography achievement data collected at time 2 only for both countries.

Perceptions of intelligence and socioeconomic status. ANOVA results for perceptions of intelligence and socioeconomic status by country at time 1 are presented in Table 3.6. Descriptive statistics are presented in Appendix 1 (Table 1.8). The results show for all measures, no significant effect of country following a multiple testing correction of $p \leq .001$ ($p = .05/90$). Results from Levene's tests showed that equal variance could be assumed for all these analyses (see Appendix 3, Table 3.7).

Table 3.6. ANOVA results for perceptions of intelligence academic and socio-economic status by country

Construct	df	*F*	*p*	η_p^2
Theories of intelligence	1,491	.006	.941	.000
Perceptions of academic and socioeconomic status	1,486	.264	.608	.001
Self-perceptions of school respect	1,466	.128	.720	.000
Self-perceptions of school grades	1,468	.613	.434	.001
Self-perceptions of family SES, occupation	1,456	.010	.921	.000
Self-perceptions of family SES, education	1,459	.210	.647	.000

Significant results in bold following a Bonferroni multiple testing correction of $p \leq .001$ ($p = .05$ divided by number of measures (k=90) across maths and geography, time 1, 2 & 3). All measures collected at time 1 only for both countries.

Discussion

The aim of the present study was to investigate potential variation in academic outcomes between two samples of 11 to 12 year old students from two countries with different education systems, Russia and the UK. The results showed, for the majority of measures, no significant mean differences between the samples across the three assessment points for maths performance, maths and geography classroom environment, motivation and subject anxiety. The only observed difference was small (5%) whereby on average, geography performance was significantly better for the UK students compared to the Russian sample across the assessment waves. For the majority of measures, variances were also equal across samples apart from the number line task at time 2 and time 3, maths homework behaviour across the assessment waves, and geography performance at time 1 and time 3. Therefore, caution should be advised when interpreting findings for geography performance as greater variance was seen in the UK compared to the Russian sample. The results showed no significant differences within and between countries across the academic year for almost all outcomes. These findings suggest that, apart from geography performance, expected differences of worse results for Russian children at time 1 compared to the UK following a lengthy summer break were not observed. Furthermore, as results did not change significantly across the summer break (between time 2 and time 3) within either sample, no impact was shown for any length of break.

The results also suggest that primarily, the UK sample is representative of the UK population as it is comparable with the large representative sample of around 8,000 UK twin pairs (TEDS). The mean scores found in the UK sample for maths self-perceived ability at time 1 (whole UK sample: M = 0.18, SD = 0.98), are highly similar to those found for 3,885 individuals in TEDS at age 12 (males, M = 0.10, SD = 1.03; females, M = - 0.08, SD = 0.97) (Kovas et al., 2015). When comparing the UK sample's average school maths achievement, it is slightly higher than the TEDS' average grades (M=4.39, SD = 0.91, N = 2577) (Luo, Haworth, & Plomin, 2010). Average grades in the UK sample fall between 5b/5a (M = 14.77, SD = 2.92, where the scale 1-7a, b, & c was recoded to 1-21). The UK sample is also slightly above the 4b that was expected in national achievement levels at the time of the study (Middlemass, 2014). The slightly higher average grades may be due to higher scores from children in one UK school who previously attended private primary education. The Russian sample's average grade in school maths achievement is 3.84 (SD

= 0.65) on a scale of 1-5 where 5 is ‘excellent’. No information was available on Russian national averages to directly compare but the score being between ‘satisfactory’ and ‘good’ suggests the sample are likely to be around average and therefore representative of the Russian population (NICARM, 2016).

The results demonstrating no significant difference between countries for perceptions of intelligence and socioeconomic status at time 1 are unexpected. As the UK is a higher SES country, a lower evaluation in the Russian sample might be anticipated. However, perceived SES is relative within the population and therefore, it is perhaps unsurprising that the results are similar. Some effect of absolute SES could be expected. For example, at lower absolute levels of SES, children with lower SES may feel particularly disadvantaged in comparison to their peers. The study did not find any such trends, perhaps because the countries are not so different in this respect. Indeed, both schools are from international cities with ample opportunity for cultural activities which have been shown to positively associate with academic outcomes (e.g. Xu & Hampden-Thompson, 2012). This resemblance in availability of cultural activities between the school regions may also contribute towards the academic similarity between them. Therefore off-setting any differences in SES suggested to impact variation in achievement outcomes (e.g. Nye, Konstantopoulos, & Hedges, 2004).

These findings are in line with PISA results that showed similar rankings and highly similar mean scores for the UK and Russia in maths performance (PISA, 2009; 2012; 2016). The small difference in geography performance between the UK and Russian students also reflects the slightly lower ranking shown for Russia compared to the UK in the PISA science results (PISA, 2016). This finding is likely due to variation in geography curricula between the two countries rather than different education systems as the two samples do not differ on any other measures.

The results also correspond with previous research that showed similarity between countries (including the UK and Russia) in self-perceived ability and enjoyment of maths (Kovas et al., 2015). The findings also offer support for TIMMS results that show England and Russia in largely similar rankings for maths performance as well as maths confidence (TIMMS, 2011; 2016). The findings for geography performance are not in line with the TIMMS results as similar rankings were found across the UK and Russia in science at these ages (TIMMS, 2011; 2016).

As the samples appear to be representative of their countries, the results imply that differences in the two education systems do not lead to differences in the majority of academic outcomes. This means that it may not be important whether or not classes and/or schools are streamed by ability. As overall, the two systems lead to very similar outcomes, despite the absence of tracking in the Russian schools. It might be suggested, however, that in the Russian school there is a form of implicit selection. By having the opportunity to learn up to two second languages, parents have elected to enroll their child into a more challenging programme and therefore, have confidence in their child's ability to succeed in this. Hence, the schools may be highly similar across both samples and therefore not subject to differences in tracking that may influence variation in achievement (Woessman, 2016). It may be that despite different education policies between the countries, the students themselves are highly similar, perhaps because both are from mixed ability schools. The UK students are only streamed by ability for their maths classes and are not from schools which restrict their intake to high ability students. The findings suggest that on average, it may not matter whether students are taught alongside students of similar ability.

In terms of their rankings in PISA and TIMMS results, neither country are at extreme ends of the distribution for maths achievement. Although there are mean differences between participating countries in the world-wide assessments for maths, reading and science, most variation is within countries. Further analysis in Chapter 4, Chapter 5 and Chapter 6 will investigate within countries, within schools and between teacher and classroom groups.

Limitations

This study has a number of limitations. Primarily, the timing of the first data collection in Russia at the start of the spring term meant that initial baseline measures when students began their academic year in September were unavailable. Likewise data were not collected from both samples of students during primary school, apart from their school achievement for maths. This meant the study was unable to assess any fluctuation in motivation across the transition period into secondary education. The study was also unable to control for participant fatigue in having to repeatedly answer the same questions at each assessment. It might also be suggested that with such a stringent multiple testing correction there is a risk of Type II error. However, only two comparisons, geography performance and maths classroom environment, revealed *p* values below 0.05 for a main effect of country (p <.001 and p =.032, respectively). This suggests that the similarity between the two samples is quite robust.

Conclusion

In conclusion, the study found no significant differences between the UK and the Russian samples for the majority of academic outcomes across one academic year in secondary education, despite the different education systems. The significant effect of country found for geography performance was small and may reflect differences in curricula between the two samples. The results were largely similar for maths and geography and reflect previous findings in mathematics and science in much larger comparisons. These findings also suggest that the samples are representative of their countries' populations. The resemblance between the two samples may result from informal selection processes in the Russian school. This similarity across samples provides a good basis from which to make further within group comparisons. These findings suggest that the two education systems lead to similar educational outcomes, and that factors that drive individual differences within populations are likely to be similar in the UK and Russia.

Chapter 4

Teacher/classroom effects

Abstract

Research investigating teacher and classroom effects on achievement has yielded modest effect sizes (Nye et al., 2004). Very little research is available for teacher/classroom effects on other outcomes, such as motivation, anxiety, peer and teacher relations. This study investigates the teacher/classroom effects on a range of outcomes, including achievement, performance, motivation, peer and teacher relations, attitudes towards the subject, and subject anxiety.

The study used a sample of 11 classes of 10-12 year old students (5th graders) in Russia. The students remain in the same class groups for their entire school education, with each group having the same primary school teacher for four years. It is therefore reasonable to expect significant average differences across these classes in all educationally relevant outcomes. The results showed no significant effects for most measures. However, a moderate effect of classroom was observed for maths and geography achievement, maths performance, classroom environment, student-teacher relations and classroom atmosphere. In separate analysis, a modest effect of subject teacher was shown across the same measures. ‘Teacher/classroom effects’ in this study refer to statistical significance of the comparison of the groups by current subject teacher. This, however, does not mean actual effect, as the results may be confounded by other factors, such as prior class achievement. Overall, these

findings suggest a weak effect of subject teacher, confounded by multiple factors, many of which stem from primary school.

Introduction

Research investigating teacher/classroom effects on school achievement has shown small effect sizes, with average effects of 8% (e.g. Nye et al., 2004). Several of these studies have used a large-scale approach, whereby data collected across school districts for administration purposes were used. The data usually consists of demographic information and school data such as grades and teacher employment records. Consequently, these studies can only investigate simple relationships, for example, average achievement gains across and within cohorts of students. Other studies have demonstrated the importance of classroom environments, such as classroom emotional climate (Reyes, Brackett, Rivers, White, & Salovey, 2012) and peer influence (Burke & Sass, 2008; Haworth, Davis, Hanscombe, Kovas, Dale, & Plomin, 2013) in relation to academic achievement.

One classroom in Russia contains on average, eighteen students, providing a rich environment for diverse peer-peer relations and teacher-student relations. The class atmosphere is a product of many interacting dynamic factors, including individual students' academic and behavioural attributes; teacher characteristics; school ethos; family backgrounds; educational policies. Research that went some way to investigate the complex nature of the classroom environment investigated the inter-relatedness of teacher-student/peer-peer relations in 713 US elementary school students aged 8 to 10 years (3rd and 4th grades) (Hughes, Im, & Wehrly, 2014). The study found that students with a reputation for a good student-teacher relationship (peer-nominated) were shown to have higher levels of academic respect (18%). In other words, a reputation of good student-teacher relationship explained 18% of the variance in status among peers for academic competence. It also explained higher levels of: acceptance by their peers (11%); and teacher-rated behavioural engagement (4%). Students' Year 4 achievement was also found to moderate the student–teacher relationship in association with academic respect (5%); and peer acceptance (3%). Additionally, the study found that a good student-teacher relationship not only predicted academic respect more greatly for higher achieving than lower achieving students, it also protected lower achieving students from lower levels of peer acceptance. Hughes and colleagues also investigated the distribution of teacher support across the classroom. They found that if a class perceived the teacher as showing preference towards a few specific students, this inequality negatively predicted peers' academic reputations (14%). The study also found that this was moderated (6%) by Year 4 achievement. These results translate as peers perceiving lower achieving students as being less competent academically than they would in a classroom where teacher support was perceived as being allocated more uniformly across peers.

Another study investigated the relationship between support and school engagement for adolescent US students (Wang & Eccles, 2012). It is argued that school engagement declines for adolescents (e.g. Wang & Holcombe, 2010), but this decline may be slower within a more supportive school environment (e.g. Eccles et al., 1993). The study found that an increase of 1 SD in teacher social support led to a lower rate of decline (0.37 SD) in school compliance. However, school compliance also decreased faster (0.28 SD) in relation to a 1 SD increase in peer support. Peers aligning with either pro- or anti-social values had positive and negative associations (respectively) with individuals' school compliance.

Much of the research into classrooms has been conducted in the US where there is a potential confounding of peer/teacher effects by streaming and tracking processes. For example, if students are assigned to classrooms based on their level of ability and this is not accounted for when comparing average classroom achievement scores, then between-classroom differences in peer/teacher effects will largely be due to differences in ability rather than actual differences in teacher or peer characteristics. Research in mixed ability classrooms in North China however, has also found effects of peer achievement in relation to individual student achievement (Carmen & Zhang, 2012). The students, aged 12 to 16 years, were in middle school (grades 7 to 9) where the school policy prescribes large mixed ability classrooms with student numbers ranging from 51 to 65. The balance of ability, however, is not left to chance by random allocation of students to classrooms; instead, at admission, students are tested in maths and Chinese to produce a total ability score. The classes are then formed to include students from all ability levels so that average ability is strictly maintained across all classes; gender balance is also preserved across classes. The students remain in these class groups for the entire three years of their middle school education. During this period, students' subject teachers also remain with them. The only change is the seating within classrooms whereby students are arranged so the tallest students are at the back and shortest at the front. Any student who has a growth spurt will be moved towards the rear of the classroom. Students are compelled to achieve as successful students are highly respected among their classmates, friends and families. In the participating school this pressure is sustained by a parent-teacher meeting held following each final exam where a spreadsheet of students' final scores and ranking is posted on the classroom wall for everyone to see. The study found that an increase in classroom peers' average maths test scores by 0.10 standard deviation (SD) led to an individual's test score increase of 0.037 SD. For Chinese, a 0.10 SD increase for peers equaled an increase of 0.042 SD for the individual. However, the effect was only found for average ability students in the classroom; students at the upper and lower end of the ability distribution were unaffected. No significant effect of peers was found for English test scores.

These studies highlight the importance of factors within the classroom environment in relation to academic outcomes. Demonstrating the impact that peers exert on school engagement and achievement. They also highlight the value of good student-teacher relations within the classroom for student engagement and academic status. By illustrating the dynamic nature of the school classroom in streamed and mixed ability classrooms, these studies emphasise the need to consider such factors when exploring teacher/class effects.

The Current Study

The current study focuses on data collected at time 1 from a longitudinal study where three assessments were made across one academic year. It utilizes data from two Russian schools that allows us the unique opportunity to explore teacher/classroom effects in a sample where students are not streamed for ability and remain in the same class groups throughout their school education. Both primary and secondary school education in Russia are conducted within one building.

Classrooms in Russia are comprised of mixed ability students and are formed when students enter primary education at age 7 years. For the first four years (primary education), all subjects, (with the exception of second language and physical education) are taught by one teacher and this same teacher remains with each class until they start secondary education. At this stage (around 11 to 12 years, 5^{th} grade), students now have specific subject teachers for the first time. Not only do students have different lessons with different teachers, but as the number of subject teachers is fewer than the number of class groups, most students' teachers teach more than one class group per subject. Anecdotally, secondary education teachers report that the primary school teacher exerts considerable influence on the classroom ethos, which persists throughout secondary education. The class group itself may also have a strong dynamic depending on the interplay of student factors such as ability, motivation, and behavior.

The sample in the current study is from two specialized linguistic schools with enhanced language curricula. In one school, students learn English and Chinese, and in the other, students learn English and/or Spanish. The classrooms are mixed ability and take children from the locality. However, enrollment into classes that offer two rather than one foreign language is not completely random, as parents have elected to enroll their children into a specialized programme that is likely more demanding than in a non-specialized school. Previous research, which included a different cohort of children from the same 2 schools used in the present study, found no average differences in achievement and cognitive performance between groups of 2nd grade children after 1 year of studying different second languages (English, Japanese, Chinese, Spanish) (Rodic, Zhou, Tikhomirova, Wei, Malykh, Ismatulina, Sabirova, Davidova, Tosto, Lemelin, & Kovas, 2015). The study also demonstrated no differences in a range of cognitive abilities between the two schools.

The aim of this study was to test whether being in the same classroom with the same peers during primary and secondary education would lead to a significant effect of teacher/classroom on measures of school achievement, performance, classroom environment, motivation and subject anxiety. To address the potential effect of the teacher, the study examines teacher/classroom effects within two separate domains, maths and geography. The following research questions are addressed: 1) Does being among the same peers for the previous four years and remaining within the same class, have an overriding influence beyond that of the current subject teacher? 2) Does having the same primary school teacher for the previous four years influence the classroom environment beyond the current subject teacher? 3) Are teacher/classroom effects similar across different domains, i.e. maths and geography?

Methods

Participants

Participants were 229 (102 males; 127 females) 10 to 12 year old students (mean age 139 months, range 127-148 months) from two urban mixed ability schools in St. Petersburg, Russia. In one school, identified here as school 1, there were eight classes of students who learned English (*n*=50) or English and Spanish (*n*=136). In the other school (school 2), there were three classes of students who learned English and Chinese (*n*=43). The students were allocated to a class when they started school at age 7 years depending on their choice of second language. They remained in these groups for their entire education across all subjects. Now in their first year of secondary education, the students have specific subject teachers for the first time. Across the two schools, the eleven classes were taught by six maths teachers and five geography teachers (see Table 4.1). Although the teachers covered more than one class, the students only had one teacher for maths or geography.

Measures

A detailed description of the measures used in this study is provided in the methods section in Chapter 2, pages 61 to 70.

Procedure

Students. Participant consent was obtained via an opt-out form that was sent home to each student's parent/guardian. Those not wishing their child to participate returned the form to exclude them from the study. Verbal consent was obtained from participants at the beginning of each data collection, and all participants were given the right to withdraw from the study at any time. Confidentiality of all participants' responses was also ensured.

Participants took part as a class exercise during their mathematics lessons under test conditions. Data were collected at three assessment points, the first at the beginning of the spring term, the second in April/May at the end of the school year, and the third in September when students returned from their summer break. At each assessment, up to two classes were tested per day so data collection took place over the course of two weeks.

After standardised instructions were read to the class, participants were presented with a range of tasks and self-report questionnaires in pencil and paper format. The first task to be presented was the Maths Problem Verification task (MPVT), which is a timed test. Eight minutes were allowed for completion of the task, following this, papers were collected to prevent participants returning to unfinished items. The participants were given the remainder of the lesson to complete the rest of the activities.

The non-cognitive measures were grouped and presented separately for each subject. Participants were asked to think about their maths classrooms since the beginning of term for the first eight measures, and asked to think about their geography classrooms for the last eight measures.

Teachers. When data were collected from students at the first assessment, their teachers were also administered a battery of measures for further analyses planned. Written informed consent was obtained prior to completion of the self-report questionnaires. Confidentiality of their responses was assured and they were given the right to withdraw from the study at any time.

Table 4.1. Class groups with their language specialism and maths and geography teachers

	C1e	C2e	C3e	C4se	C5se	C7se	C8se	C6se	C9ce	C10ce	C11ce
N	23	9	18	28	25	24	31	28	18	11	14
Second Language	E	E	E	E & S	E & S	E & S	E & S	E & S	E & C	E & C	E & C
Maths Teacher	TM3	TM3	TM4	TM6	TM6	TM6	TM6	TM5	TM1	TM2	TM2
Geography Teacher	TG4	TG4	TG5	TG5	TG2	TG2	TG2	TG3	TG1	TG1	TG1

E = English; E & S = English and Spanish; E & C = English and Chinese. Class groups are identified by number (1-11) and language specialism: e = English; se = English & Spanish; ce = English & Chinese.

Analyses

Analyses were conducted using data collected at the first assessment (time 1) on variables corrected for age, with outliers (± 3SD) removed. A Bonferroni multiple testing correction was set of $p \leq .001$ where $p = .05$ was divided by the number of measures (k=70) across the two schools and across maths and geography at time 1. This translates as: maths classroom measures = 14 x 2 (14 measures assessed separately within school 1 and school 2); geography classroom measures = 13 x 2 (13 measures assessed separately within school 1 and school 2); maths achievement =1 x 2 (1 measure assessed separately within school 1 and school 2); geography achievement = 1 x 2 (1 measure assessed separately within school 1 and school 2); perceptions of intelligence and socioeconomic status = 6 x 2 (6 measures assessed separately within school 1 and school 2).

Analyses of variance (ANOVA) were conducted within each school to assess potential differences in means for school achievement, performance, classroom environment, classroom atmosphere, motivation, attitudes towards subject and subject anxiety by classroom at time 1. Planned pairwise comparisons were also conducted between classrooms applying a Dunnett's T3 multiple comparison correction as it maintains tight control of the Type 1 error rate while allowing for differences in variances and group size (Field, 2011).

To further investigate potential effects of the teacher/classroom, students' classes were regrouped to account for secondary school teachers teaching more than one class. To differentiate between primary school teachers and other current class teachers, current maths or geography teachers will herein be termed as ‘subject teachers’, unless otherwise mentioned specifically as maths/geography teachers.

With teacher groups combined across the two schools, additional ANOVAs were conducted by teacher group for all the measures within each domain. For these analyses a Bonferroni multiple testing correction was set of $p \leq .001$ where $p = .05$ divided by the number of measures (k=41) across the two sets of teachers (maths and geography) at time 1. This translates as: maths classroom measures = 14 x 1 (14 measures assessed for maths teacher groups); geography classroom measures = 13 x 1 (13 measures assessed for geography teacher groups); maths achievement = 1 x 1 (1 measure assessed for maths teacher groups); geography achievement = 1 x 1 (1 measure assessed for geography teacher groups); perceptions of intelligence and socioeconomic status = 6 x 2 (6 measures assessed separately for maths and geography teacher groups). Planned pairwise comparisons were also conducted between the teacher groups, using a Dunnett’s T3 multiple comparison correction.

Further analyses were conducted to establish any differences in ability associated with learning more than one second language. Primary school achievement was selected to explore potential implicit selection processes that might be linked with this particular measure. ANOVA were conducted on maths and geography primary school achievement by linguistic specialism. For these analyses a Bonferroni multiple testing correction was set of $p \leq .025$ where $p =$ .05 divided by the number of measures (k=2) across maths and geography at time 1.

Following the observation of any significant differences between classes/teachers for any measures, the classes and teacher groups were ranked by means, highest to lowest, to assess any correspondence of class ranking across the significant measures within each domain. To establish any correspondence of ranking between domains, class ranks were compared between maths and geography measures. These rankings were then compared with primary school achievement to examine any similarity of ranking across primary and secondary teachers/classes. To ascertain any influence of teacher, bivariate correlations were conducted to investigate the strength of the relationships between corresponding measures across the two domains.

Results

Means, standard deviations (SD) and N for all assessed variables can be found in Tables 4.2 to 4.7 by classroom, in Tables 4.11 to 4.14 by teacher and Table 4.18 by linguistic specialism.

Class groups are identified by a number (1 to 11) prefixed by 'C' to distinguish from teacher groups; language specialism is indicated as follows: e = English; se = English & Spanish; ce = English & Chinese. Maths teacher groups are also identified by a number (1 to 6) prefixed by 'T' and 'M' to distinguish from geography teachers; geography teachers are identified by a number (1 to 5) prefixed by 'TG'.

Differences Between Maths Classrooms

School 1

ANOVA results by maths classroom are presented in Table 4.2 for school 1. The results show that for the majority of measures, there was no significant effect of maths classroom following multiple testing correction of $p \leq$

.001 (p = .05/70). Levene's tests revealed equal variance for most measures except maths primary school achievement, maths performance, maths classroom environment and student-teacher relations (see Appendix 4, Table 4.1). A significant effect of classroom was found for the following four measures:

Maths primary school achievement. A moderate effect of classroom was found, $F(7,164) = 7.341$, $p < .001$, $\eta_p^2 = .239$, with the highest mean score revealed for class C8se and the lowest for C3e. Pairwise comparisons also showed that C8se had significantly higher primary school achievement than C1e ($p = .001$), C5se ($p < .001$), and C3e ($p < .001$), following multiple testing correction of $p \leq .001$ ($p = .05/70$). However, Levene's test revealed unequal variances for these analyses ($p = .05$), with the smallest variance shown for class C7se (0.55) and the largest for class C1e (1.10).

Maths performance. A moderate effect of classroom was observed, $F(7,178) = 9.147$, $p < .001$, $\eta_p^2 = .265$, with the highest mean score shown for C6se and the lowest for C5se. Only three pairwise comparisons reached significance following multiple testing correction of $p \leq .001$ ($p = .05/70$): C5se had significantly lower maths performance than C6se ($p < .001$), C7se ($p < .001$), and C8se ($p < .001$). However, Levene's test revealed unequal variances for these analyses ($p = .03$), with the smallest variance shown for class C5se (0.44) and the largest for two classes, C1e and C2e (1.37).

Student-teacher relations. A modest effect of classroom was found, $F(7,176) = 3.699$, $p = .001$, $\eta_p^2 = .128$, with class C6se showing the highest mean score and class C7se showing the lowest. No pairwise comparisons reached significance following multiple testing correction of $p \leq .001$ ($p = .05/70$). However, Levene's test revealed unequal variances for these analyses ($p = .01$), with the smallest variance shown for class C6se (0.34) and the largest for class C4se (1.34).

Maths classroom chaos. A modest effect of classroom was revealed, $F(7,176) = 4.087$, $p < .001$, $\eta_p^2 = .140$, with the highest mean score (low chaos) shown for C2e and the lowest for C4se (high chaos). However, pairwise comparisons showed that C4se only had significantly higher levels of chaos than C6se ($p < .001$), following multiple testing correction of $p \leq .001$ ($p = .05/70$). Levene's test revealed equal variances were assumed for these analyses ($p = .58$). While it appears unusual that the difference between highest (C2e) and lowest (C4se) means did not reach significance despite having the largest mean difference of 1.24, this pairwise comparison also had the largest standard error of 0.34. This is compared to the mean difference and standard error between C4se and C6se of 1.16 (*SE* = 0.23).

Table 4.2. Maths classroom variables for school 1: Means, standard deviation (SD), and N with ANOVA results by classroom

Construct	C1e	C2e	C3e	C4se	C5se	C6se	C7se	C8se	*p*	η_p^2
Maths primary school achievement	-0.43 (1.05) n=21	0.01 (1.01) n=9	-0.63 (0.79) n=18	-0.02 (0.77) n=27	-0.36 (1.01) n=23	0.48 (0.87) n=28	0.06 (0.74) n=23	0.91 (0.76) n=23	**.000**	**.239**
Maths performance	-0.08 (1.17) n=23	0.40 (1.17) n=9	-0.34 (1.03) n=18	-0.27 (0.86) n=28	-1.15 (0.66) n=25	0.69 (0.80) n=28	-0.21 (0.68) n=24	0.01 (0.81) n=31	**.000**	**.265**
Number line	0.11 (1.03) n=22	0.01 (1.00) n=9	0.24 (0.71) n=18	-0.17 (0.76) n=28	0.08 (0.75) n=25	-0.42 (1.17) n=28	0.07 (0.76) n=23	0.13 (0.80) n=30	.196	.054
Maths self-perceived ability	0.09 (1.09) n=23	-0.01 (0.72) n=9	-0.01 (1.11) n=17	-0.07 (0.96) n=28	-0.34 (0.87) n=24	0.18 (0.88) n=28	0.45 (0.83) n=24	0.08 (1.02) n=28	.236	.051
Maths enjoyment	0.20 (0.84) n=22	0.52 (0.67) n=9	0.22 (0.80) n=17	-0.26 (1.07) n=24	-0.04 (0.81) n=21	0.06 (0.89) n=28	0.14 (0.83) n=24	0.15 (0.87) n=28	.398	.043
Maths classroom environment	-0.16 (1.03) n=23	0.52 (0.62) n=9	0.20 (0.86) n=17	0.04 (1.19) n=27	0.11 (0.76) n=25	0.61 (0.60) n=28	-0.35 (1.09) n=24	0.22 (0.78) n=31	.010	.098
Maths classroom Student-teacher relations	-0.25 (1.03) n=23	0.57 (0.69) n=9	0.23 (0.79) n=17	-0.02 (1.16) n=27	-0.04 (0.83) n=25	0.72 (0.58) n=28	-0.29 (1.00) n=24	0.20 (0.80) n=31	**.001**	**.128**
Maths classroom peer competition	-0.11 (0.96) n=23	0.33 (0.77) n=8	0.03 (0.81) n=18	0.11 (0.97) n=27	0.55 (0.67) n=25	-0.11 (1.15) n=28	-0.38 (1.09) n=24	0.15 (0.83) n=31	.042	.078

Significant results in bold following a Bonferroni multiple testing correction of $p \leq .001$ where $p = .05$ divided by the number of measures (k=70) across maths and geography at time 1 and across the two schools. Classes learning: English= C1e; C2e; C3e; English and Spanish= C4se; C5se; C6se; C7se; C8se.

Table 4.2. Continued. Maths classroom variables for school 1: Means, standard deviation (SD), and N with ANOVA results by classroom

Construct	C1e	C2e	C3e	C4se	C5se	C6se	C7se	C8se	*p*	η_p^2
Maths classroom chaos	-0.21 (0.92) n=23	0.65 (0.87) n=9	0.02 (0.96) n=18	-0.59 (0.97) n=28	0.05 (1.03) n=25	0.58 (0.78) n=28	0.11 (1.04) n=24	-0.29 (1.04) n=29	**.000**	**.140**
Maths homework behaviour	-0.10 (1.04) n=23	0.11 (0.83) n=9	0.51 (1.01) n=18	-0.14 (1.09) n=28	0.14 (1.09) n=25	-0.51 (1.04) n=28	-0.33 (1.08) n=24	0.00 (0.98) n=29	.065	.072
Maths homework feedback	-0.42 (0.99) n=23	0.46 (0.83) n=9	-0.28 (1.35) n=18	0.03 (1.09) n=28	0.32 (1.10) n=25	0.14 (0.76) n=28	-0.10 (0.93) n=24	0.03 (0.97) n=29	.166	.057
Maths homework total scale	-0.26 (0.86) n=23	0.36 (0.90) n=9	-0.49 (1.27) n=18	0.05 (1.02) n=28	0.31 (0.92) n=24	0.30 (0.87) n=28	0.07 (0.95) n=24	0.02 (0.88) n=29	.085	.068
Maths environment	0.07 (1.07) n=21	0.49 (1.06) n=9	-0.16 (1.05) n=17	-0.02 (1.11) n=27	0.05 (0.89) n=25	0.58 (0.69) n=28	0.04 (0.92) n=24	-0.41 (0.87) n=28	.013	.098
Maths usefulness	0.02 (0.82) n=22	-0.41 (0.61) n=9	0.07 (1.06) n=17	-0.03 (0.94) n=28	-0.04 (0.85) n=23	-0.41 (0.74) n=28	-0.10 (1.03) n=23	0.08 (1.14) n=28	.528	.035
Maths anxiety	-0.24 (1.03) n=23	-0.18 (0.73) n=9	-0.61 (0.78) n=17	0.12 (1.01) n=26	0.59 (0.81) n=24	0.00 (1.06) n=28	-0.02 (0.95) n=24	-0.02 (0.97) n=28	.012	.098

Significant results in bold following a Bonferroni multiple testing correction of $p \leq .001$ where $p = .05$ divided by the number of measures (k=70) across maths and geography at time 1 and across the two schools. Classes learning: English= C1e; C2e; C3e; English and Spanish= C4se; C5se; C6se; C7se; C8se.

Table 4.3. Maths classroom variables for school 2: Means, standard deviation (SD), and N with ANOVA results by classroom

Construct	C9ce	C10ce	C11ce	*p*	η_p^2	Construct	C9ce	C10ce	C11ce	*p*	η_p^2
Maths primary school achievement	0.29 (0.96) n=16	-0.12 (1.08) n=10	-0.39 (0.84) n=9	.230	.088						
Maths performance	0.66 (0.83) n=18	0.56 (0.72) n=11	0.59 (0.56) n=14	.921	.004	Maths classroom chaos	0.27 (0.71) n=18	0.69 (1.01) n=11	-0.64 (0.54) n=14	**.000**	**.347**
Number line	-0.90 (1.19) n=18	0.21 (0.43) n=11	-0.34 (0.68) n=14	.009	.211	Maths homework behaviour	-0.17 (0.85) n=18	0.40 (0.70) n=11	0.38 (1.23) n=14	.186	.081
Maths self-perceived ability	-0.22 (0.99) n=16	-0.16 (0.92) n=9	-0.10 (1.17) n=13	.952	.003	Maths homework feedback	0.22 (0.87) n=18	0.10 (0.69) n=11	-0.42 (0.97) n=14	.114	.103
Maths enjoyment	0.15 (1.02) n=15	-0.19 (0.70) n=9	0.01 (0.98) n=13	.686	.022	Maths homework total scale	0.27 (0.65) n=18	-0.06 (0.79) n=11	-0.58 (1.10) n=14	.028	.164
Maths classroom environment	-0.63 (0.80) n=17	-0.02 (0.67) n=9	-0.46 (0.95) n=13	.218	.081	Maths environment	-0.27 (1.15) n=17	-0.19 (0.98) n=9	-0.25 (1.12) n=14	.986	.001
Maths classroom student-teacher relations	-0.60 (0.94) n=17	0.14 (0.68) n=9	-0.44 (0.93) n=13	.133	.106	Maths usefulness	0.13 (0.68) n=18	-0.11 (0.82) n=10	0.26 (1.00) n=14	.552	.030
Maths classroom peer competition	-0.04 (0.95) n=17	-0.50 (0.43) n=9	-0.06 (1.29) n=13	.501	.038	Maths anxiety	0.07 (0.95) n=17	-0.65 (0.60) n=10	0.22 (1.08) n=14	.072	.129

Significant results in bold following a Bonferroni multiple testing correction of $p \leq .001$ where $p = .05$ divided by the number of measures (k=70) across maths and geography at time 1 and across the two schools. All classes learning English and Chinese.

School 2

ANOVA results for school 2 by maths classroom are presented in Table 4.3 and show no significant effect of maths classroom for all of the measures apart from classroom chaos, following multiple testing correction of $p \leq .001$ ($p = .05/70$). Levene's tests revealed equal variance for most measures except peer competition (see Appendix 4, Table 4.2).

Maths classroom chaos. A moderate effect of classroom was observed, $F(2,40) = 10.628$, $p < .001$, $\eta_p^2 = .347$, with the highest mean score (low chaos) shown for C10ce and the lowest (high chaos) for C11ce. Pairwise comparisons showed C11ce had significantly higher levels of chaos than C9ce ($p = .001$), following multiple testing correction of $p \leq .001$ ($p = .05/70$). Levene's test revealed equal variances were assumed for these analyses ($p = .51$). Similarly to school one for this measure, the absence of effect between the highest and lowest means is likely due to the larger standard error between C10ce and C11ce despite having the largest mean difference of 1.33 (SE = 0.33). This is compared to the mean difference and standard error between C11ce and C9ce of .91 (SE = 0.22). Further, the pairwise comparison between C10ce and C11ce did not survive the stringent multiple testing correction (p = .004).

Differences Between Geography Classrooms

School 1

ANOVA results for school 1 by geography classroom are presented in Tables 4.4. Similarly to maths, the results show for the majority of measures, no significant effect of geography classroom following multiple testing correction of $p \leq .001$ ($p = .05/70$). Levene's tests revealed equal variance for most measures except geography classroom environment and student-teacher relations (see Appendix 4, Table 4.3). A significant effect of classroom was found for the following five measures:

Geography primary school achievement. A moderate effect of classroom was revealed, $F(7,165) = 7.681$, $p < .001$, $\eta_p^2 = .246$, with the highest mean score shown for C6se and the lowest for C3e. Pairwise comparisons showed that C6se had significantly higher primary school achievement than C1e ($p < .001$). C3e had significantly lower achievement than C6se ($p < .001$), C4se ($p < .001$), and C8se ($p < .001$). Levene's test revealed equal variances were assumed for these analyses ($p = .33$).

Geography classroom environment. A modest effect of classroom was observed, $F(7,166) = 4.805$, $p < .001$, $\eta_p^2 = .168$, with the highest mean score shown for C6se and the lowest for C8se. Pairwise comparisons showed C6se rated their classroom environment significantly higher than C1e ($p < .001$), and C8se ($p < .001$). However, Levene's test revealed unequal variances for these analyses ($p = .003$), with the smallest variance shown for class C6se (0.24) and the largest for class C4se (1.19).

Student-teacher relations. A moderate effect of classroom was shown, $F(7,166) = 5.544$, $p < .001$, $\eta_p^2 = .189$, with the highest mean score observed for C6se and the lowest for C1e. Pairwise comparisons showed C6se rated their student-teacher relationship significantly higher than C1e ($p < .001$), C8se ($p < .001$), and C4se ($p < .001$), following multiple testing correction of $p \leq .001$ ($p = .05/70$). However, Levene's test revealed unequal variances for these analyses ($p = .02$), with the smallest variance shown for class C6se (0.31) and the largest for class C7se (1.37).

Geography classroom chaos. A moderate effect of classroom was revealed, $F(7,168) = 5.043$, $p < .001$, $\eta_p^2 = .174$, with the highest mean score (low chaos) shown for C2e and the lowest score (high chaos) for C5se. The only significant pairwise comparison following multiple testing correction of $p \leq .001$ ($p = .05/70$) showed that C5se only had significantly higher levels of chaos than C6se ($p = .001$). Levene's test revealed equal variances were assumed for these analyses ($p = .39$). Similarly to maths classroom chaos, the difference between the highest (C2e) and lowest (C5se) means did not reach significance despite having the largest mean difference of -1.37. This pairwise comparison, however, also had the largest standard error of 0.33. This is compared to the mean difference and standard error between C5se and C6se of -1.11 (*SE* = 0.23). Further, the pairwise comparison between C2e and C5se did not survive the stringent multiple testing correction ($p = .013$).

Geography environment. A moderate effect of classroom was observed, $F(7,160) = 4.869$, $p < .001$, $\eta_p^2 = .176$, with the highest mean score shown for C6se and the lowest score for C1e. This was the only significant pairwise comparison following multiple testing correction of $p \leq .001$ ($p = .05/70$): C6se rated their geography environment significantly higher than C1e ($p < .001$). Levene's test revealed equal variances were assumed for these analyses ($p = .65$).

School 2

ANOVA results for school 2 by geography classroom are presented in Table 4.5 and similarly to maths, show no significant effect of classroom for any of the measures following multiple testing correction of $p \leq .001$ ($p = .05/70$). Levene's tests revealed equal variance for most measures except geography classroom environment, student-teacher relations, and homework behaviour, homework total scale, and geography anxiety (see Appendix 4, Table 4.4).

Table 4.4. Geography classroom variables for school 1: Means, standard deviation (SD), and N with ANOVA results by classroom

Construct	C1e	C2e	C3e	C4se	C5se	C6se	C7se	C8se	*p*	η_p^2
Geography primary school achievement	-0.51 (0.81) n=22	0.19 (0.73) n=9	-0.96 (0.75) n=18	0.34 (0.93) n=27	0.00 (1.07) n=23	0.67 (0.80) n=28	-0.24 (1.02) n=23	0.38 (0.87) n=23	**.000**	**.246**
Geography performance	-0.20 (0.90) n=23	0.29 (0.77) n=9	0.05 (0.62) n=18	-0.03 (0.91) n=22	-0.23 (0.91) n=25	0.58 (0.68) n=28	0.41 (0.50) n=24	0.03 (0.75) n=31	.002	.124
Geography self-perceived ability	-0.02 (0.74) n=21	0.22 (0.70) n=9	0.36 (1.26) n=17	-0.10 (1.09) n=27	-0.40 (1.01) n=22	0.18 (1.03) n=26	0.16 (0.81) n=23	-0.17 (1.06) n=24	.282	.051
Geography enjoyment	0.16 (0.99) n=22	0.04 (0.57) n=8	0.17 (1.15) n=17	-0.06 (0.99) n=25	-0.18 (1.09) n=23	0.15 (1.07) n=27	-0.07 (0.99) n=22	-0.18 (0.96) n=26	.851	.020
Geography classroom environment	-0.31 (0.88) n=21	0.39 (0.63) n=9	-0.19 (1.02) n=17	-0.12 (1.06) n=26	0.33 (0.73) n=23	0.82 (0.49) n=28	-0.13 (1.09) n=23	-0.34 (1.14) n=27	**.000**	**.168**
Geography classroom Student-teacher relations	-0.41 (0.98) n=21	0.26 (0.80) n=9	-0.17 (1.17) n=17	-0.23 (1.00) n=26	0.36 (0.79) n=23	0.91 (0.56) n=28	-0.21 (1.06) n=23	-0.20 (0.99) n=27	**.000**	**.189**
Geography classroom peer competition	-0.18 (0.87) n=21	0.63 (0.69) n=9	-0.33 (1.06) n=17	0.21 (1.00) n=26	0.35 (0.63) n=23	-0.03 (0.88) n=28	0.14 (1.09) n=23	-0.15 (1.20) n=27	.130	.064

Significant results in bold following a Bonferroni multiple testing correction of $p \leq .001$ where $p = .05$ divided by the number of measures (k=70) across maths and geography at time 1 and across the two schools. Classes learning: English= C1e; C2e; C3e; English and Spanish= C4se; C5se; C6se; C7se; C8se.

Table 4.4. Continued. Geography classroom variables for school 1: Means, standard deviation (SD), and N with ANOVA results by classroom

Construct	C1e	C2e	C3e	C4se	C5se	C6se	C7se	C8se	*p*	η_p^2
Geography classroom chaos	-0.40 (1.05) n=22	0.85 (0.78) n=9	0.24 (0.92) n=17	-0.34 (1.11) n=27	-0.52 (0.89) n=23	0.59 (0.72) n=28	0.07 (0.95) n=23	-0.12 (0.88) n=27	**.000**	**.174**
Geography homework behaviour	-0.16 (1.05) n=22	0.07 (0.76) n=9	0.18 (1.02) n=17	-0.17 (1.02) n=26	0.22 (1.22) n=23	-0.27 (0.88) n=28	-0.11 (1.07) n=22	-0.05 (1.02) n=27	.716	.027
Geography homework feedback	-0.22 (0.97) n=22	0.16 (0.58) n=9	-0.05 (1.23) n=17	-0.57 (1.12) n=26	0.15 (0.93) n=23	0.39 (0.78) n=28	0.12 (1.01) n=22	-0.10 (1.00) n=27	.036	.085
Geography homework total scale	-0.12 (0.97) n=22	0.18 (0.68) n=9	-0.12 (1.22) n=17	-0.35 (1.10) n=26	-0.05 (1.10) n=23	0.44 (0.79) n=28	0.14 (1.05) n=22	-0.07 (0.89) n=27	.190	.057
Geography environment	-0.64 (1.11) n=22	-0.28 (0.77) n=6	-0.41 (1.03) n=17	-0.04 (1.01) n=26	-0.11 (0.91) n=21	0.77 (0.70) n=28	0.00 (0.86) n=22	-0.04 (0.90) n=26	**.000**	**.176**
Geography usefulness	-0.34 (1.02) n=21	-0.15 (0.82) n=9	-0.06 (1.07) n=16	-0.36 (0.83) n=25	0.01 (0.68) n=23	0.09 (1.39) n=28	0.22 (1.07) n=22	-0.15 (0.94) n=28	.515	.037
Geography anxiety	-0.50 (0.62) n=21	-0.37 (0.99) n=9	-0.43 (0.88) n=17	0.24 (1.00) n=24	0.43 (0.99) n=20	0.13 (0.81) n=28	-0.09 (0.91) n=22	-0.11 (0.89) n=28	.009	.108

Significant results in bold following a Bonferroni multiple testing correction of $p \leq .001$ where $p = .05$ divided by the number of measures (k=70) across maths and geography at time 1 and across the two schools. Classes learning: English= C1e; C2e; C3e; English and Spanish= C4se; C5se; C6se; C7se; C8se.

Table 4.5. Geography classroom variables for school 2: Means, standard deviation (SD), and N with ANOVA results by classroom

Construct	C9ce	C10ce	C11ce	*p*	η_p^2	Construct	C9ce	C10ce	C11ce	*p*	η_p^2
Geography primary school achievement	0.60 (0.73) n=16	-0.53 (1.08) n=10	-0.48 (0.70) n=9	.002	.323	Geography classroom chaos	-0.01 (0.92) n=17	0.54 (1.16) n=11	-0.15 (0.86) n=14	.197	.080
Geography performance	0.28 (0.90) n=18	0.01 (0.92) n=11	-0.18 (0.67) n=14	.308	.057	Geography homework behaviour	0.00 (0.83) n=17	0.02 (0.83) n=11	0.59 (1.12) n=14	.176	.085
Geography self-perceived ability	0.24 (0.75) n=17	-0.18 (0.75) n=11	0.04 (1.07) n=14	.464	.039	Geography homework feedback	0.26 (0.82) n=17	-0.01 (1.02) n=11	0.02 (1.07) n=14	.699	.018
Geography enjoyment	0.18 (0.87) n=17	-0.02 (0.65) n=10	-0.14 (1.31) n=14	.669	.021	Geography homework total scale	0.24 (0.69) n=17	0.05 (0.67) n=11	-0.32 (1.39) n=14	.287	.062
Geography classroom environment	-0.25 (0.86) n=16	-0.29 (0.79) n=10	-0.14 (1.31) n=14	.932	.004	Geography environment	0.15 (0.89) n=16	0.03 (0.94) n=9	0.21 (1.21) n=14	.919	.005
Geography classroom Student-teacher relations	-0.12 (0.73) n=16	-0.17 (0.78) n=10	-0.34 (1.20) n=14	.797	.012	Geography usefulness	0.49 (0.67) n=16	0.58 (0.64) n=11	0.07 (1.06) n=14	.236	.073
Geography classroom peer competition	-0.44 (0.95) n=16	-0.49 (1.11) n=10	0.27 (1.07) n=14	.118	.109	Geography anxiety	-0.17 (0.91) n=15	-0.13 (0.70) n=10	0.35 (1.25) n=13	.342	.060

Significant results in bold following a Bonferroni multiple testing correction of $p \leq .001$ where $p = .05$ divided by the number of measures (k=70) across maths and geography at time 1 and across the two schools. All classes learning English and Chinese.

Table 4.6. Perceptions of intelligence, and academic and socio-economic status variables for School 1: Means, standard deviation (SD) and N by classroom with ANOVA results for classroom

Construct	C1e	C2e	C3e	C4se	C5se	C6se	C7se	C8se	*p*	η_p^2
Theories of intelligence	-0.20 (0.93) n=22	0.59 (1.22) n=9	-0.24 (0.95) n=17	-0.17 (1.13) n=27	0.17 (1.13) n=25	0.19 (0.90) n=28	0.27 (0.87) n=23	-0.36 (0.86) n=27	.978	.001
Perceptions of academic and socio-economic status mean score	0.13 (1.01) n=23	0.06 (1.02) n=9	0.40 (1.09) n=17	-0.14 (1.22) n=27	-0.23 (0.90) n=23	0.03 (0.68) n=28	0.03 (1.07) n=23	0.13 (1.02) n=27	.332	.058
Perceptions Of School Respect	-0.11 (0.96) n=21	0.15 (0.90) n=9	0.04 (0.99) n=15	-0.13 (1.09) n=25	-0.09 (0.88) n=23	0.17 (0.93) n=26	0.02 (0.97) n=22	0.05 (1.02) n=25	.872	.008
Perceptions Of School Grades	0.12 (0.89) n=21	-0.14 (0.96) n=8	0.18 (1.07) n=17	-0.23 (1.22) n=26	-0.21 (0.98) n=23	-0.04 (1.02) n=26	0.26 (0.98) n=23	0.16 (0.90) n=26	.867	.008
Perceptions of family occupation	0.03 (0.97) n=23	-0.29 (0.92) n=8	0.76 (1.03) n=17	0.15 (0.95) n=24	-0.11 (1.04) n=22	0.18 (0.66) n=25	-0.44 (1.05) n=21	0.13 (1.04) n=25	.028	.176
Perceptions of family education	0.27 (0.81) n=23	0.30 (0.98) n=9	0.03 (0.93) n=16	0.03 (1.21) n=24	-0.29 (0.89) n=22	-0.09 (0.91) n=26	0.18 (1.06) n=21	-0.03 (1.00) n=27	.202	.090

Significant results in bold following a Bonferroni multiple testing correction of $p \leq .001$ where $p = .05$ divided by the number of measures (k=70) across maths and geography at time 1 and across the two schools. Classes learning: English= C1e; C2e; C3e; English and Spanish= C4se; C5se; C6se; C7se; C8se.

Differences Between Classrooms For Perceptions Of Intelligence, Academic And Socioeconomic Status

ANOVA results for perceptions of intelligence and academic and socioeconomic status by classroom can be seen in Table 4.6 (school 1) and Table 4.7 (school 2). No significant effects of classroom were found within the two schools for these constructs following multiple testing correction of $p \leq .001$ ($p = .05/70$). Levene's tests revealed equal variance for all measures in school 1 and most measures in school 2 except self-perceptions of school respect ($p = .020$) (see Appendix 4, Table 4.5).

Table 4.7. Perceptions of intelligence, and academic and socio-economic status variables for School 2: Means, standard deviation (SD) and N by classroom with ANOVA results for classroom

Construct	C9ce	C10ce	C11ce	*p*	η_p^2
Theories of intelligence	0.07 (0.87) n=17	0.02 (0.98) n=11	0.00 (1.15) n=14	.978	.001
Perceptions of academic and socio-economic status mean score	-0.40 (0.95) n=16	-0.15 (0.59) n=11	0.14 (1.23) n=13	.332	.058
Perceptions Of School Respect	-0.09 (1.09) n=15	-0.08 (0.75) n=10	0.13 (1.51) n=13	.872	.008
Perceptions Of School Grades	0.03 (0.87) n=16	-0.13 (0.75) n=10	-0.15 (1.24) n=13	.867	.008
Perceptions of family occupation	-0.76 (1.03) n=16	-0.09 (0.77) n=11	0.17 (0.87) n=13	.028	.176
Perceptions of family education	-0.49 (1.22) n=15	0.05 (1.00) n=10	0.23 (0.88) n=12	.202	.090

Significant results in bold following a Bonferroni multiple testing correction of $p \leq .001$ where $p = .05$ divided by the number of measures (k=70) across maths and geography at time 1 and across the two schools. All classes learning English and Chinese.

Class Ranking By Mean Score

To further examine the effect of teacher/class on the measures, the classes were ranked within schools by their mean scores, from highest to lowest, for all measures that reached significance.

Maths classroom. Table 4.8 and 4.9 show class rankings for all maths measures between the class groups, for school 1 and school 2 respectively. The results show some correspondence of rank for some classes across the study measures (maths performance, student-teacher relations and classroom chaos). For example in school 1, class C6se is in the top ranks with 1^{st} and 2^{nd} place, and class C2e also ranks higher with 2^{nd} place for 2 measures and 1^{st} for one. Classes C1e, C3e, C4se and C5se are in the lower ranks for most measures. Classes C7se and C8se show a less consistent pattern across the measures. Complete correspondence is shown between maths classroom environment and student-teacher relations for classes C2e and C6se only. The remaining classes show similarity of rank although they are not completely identical. We would expect such consistency as student-teacher relations is a subscale of maths classroom environment. In school 2, only classroom chaos showed a significant effect of classroom. A comparison of classroom chaos with primary school achievement revealed class C11ce in 3^{rd} place for both measures.

Geography classroom. The rankings for the geography measures between the classes in school 1 can be seen in Table 4.10 (no significant effects of class were seen for school 2). Similarly to the maths measures, the results show some correspondence of rank for some classes across the study measures (classroom environment, student-teacher relations, classroom chaos and geography environment). Given that student-teacher relations is a subscale of classroom environment, it is surprising that correspondence occurs across the two subscales for just one class (C6se in 1st place); four other classes show similarity of position across the two measures but are not completely consistent. Consistently in the higher ranks is class C6se in 1st and 2nd place. Class C2e is also at the higher end for the majority of measures. Classes C1e, C3e, C7se and C8se are ranked consistently at the lower end across the measures.

Overall across maths and geography, the consistency of rank appears to be similar. The results mostly show variation across all measures with some correspondence for specific classes. For example, C6se ranks at the high end and C1e ranks towards the lower end across both domains. This raises the question of how much influence originates from the subject teacher.

Table 4.8. Maths classroom variables for school 1: Classrooms ranked by means (highest = 1 to lowest = 8) for measures demonstrating a significant effect of maths classroom

Construct	C1e	C2e	C3e	C4se	C5se	C6se	C7se	C8se	*p*	η_p^2
Maths Primary school achievement	**7th**	**4th**	**8th**	**5th**	**6th**	**2nd**	**3rd**	**1st**	**.000**	**.239**
Maths performance	**4th**	**2nd**	**7th**	**6th**	**8th**	**1st**	**5th**	**3rd**	**.000**	**.265**
Maths classroom Student-teacher relations	**7th**	**2nd**	**3rd**	**5th**	**6th**	**1st**	**8th**	**4th**	**.001**	**.128**
Maths classroom chaos	**6th**	**1st**	**5th**	**8th**	**4th**	**2nd**	**3rd**	**7th**	**.000**	**.140**

Significant results in bold following a Bonferroni multiple testing correction of $p \leq .001$ where $p = .05$ divided by the number of measures (k=70) across maths and geography at time 1 and across the two schools. Classes learning: English= C1e; C2e; C3e; English and Spanish= C4se; C5se; C6se; C7se; C8se.

Table 4.9. Maths classroom variables for school 2: Classrooms ranked by means (highest = 1 to lowest = 3) for measures demonstrating a significant effect of maths classroom

Construct	C9ce	C10ce	C11ce	*p*	η_p^2
Maths Primary school achievement*	**1st**	**2nd**	**3rd**	.230	.088
Maths classroom chaos	**2nd**	**1st**	**3rd**	**.000**	**.347**

Significant results in bold following a Bonferroni multiple testing correction of $p \leq .001$ where $p = .05$ divided by the number of measures (k=70) across maths and geography at time 1 and across the two schools. All classes learning English and Chinese.*Not significant but used to make comparison with primary school.

Table 4.10. Geography classroom variables for school 1: Classrooms ranked by means (highest = 1 to lowest = 8) for measures demonstrating a significant effect of geography classroom

Construct	C1e	C2e	C3e	C4se	C5se	C6se	C7se	C8se	*p*	η_p^2
Geography primary school achievement	7th	4th	8th	3rd	5th	1st	6th	2nd	.000	.246
Geography classroom environment	7th	2nd	6th	4th	3rd	1st	5th	8th	.000	.168
Geography classroom Student-teacher relations	8th	3rd	4th	7th	2nd	1st	6th	5th	.000	.189
Geography classroom chaos	7th	1st	3rd	6th	8th	2nd	4th	5th	.000	.174
Geography environment	8th	6th	7th	4th	5th	1st	2nd	3rd	.000	.176

Significant results in bold following a Bonferroni multiple testing correction of $p \leq .001$ where $p = .05$ divided by the number of measures (k=70) across maths and geography at time 1 and across the two schools. Classes learning: English= C1e; C2e; C3e; English and Spanish= C4se; C5se; C6se; C7se; C8se.

Differences Between Teachers

To establish any influence of subject teacher, further analyses were conducted where students' classes were regrouped to account for secondary school teachers teaching more than one class. The eleven classes across the two schools were grouped by maths teacher (six teachers across eleven classes), and by geography teacher (five teachers across eleven classes). Table 4.1 shows each class and their corresponding teacher. Some teachers, for example TM6, teach several classes, while others like TM5, teach just one class of this year group. Teaching load of individual teachers is made up of classes of different year groups, so one maths teacher may teach, for example, 6 classes of the same year group, or 6 classes from different year groups.

ANOVAs were conducted by teacher to assess whether differences remained between these new groupings for the measures that demonstrated a significant effect of classroom across the two domains. Measures were tested for each set of teachers within each domain, this provided a multiple testing correction of $p \leq .001$ where $p = .05$ is divided by the number of measures (k=41). Primary school subjects (maths and geography achievement) were included in these analyses, even though they were not taught by this set of teachers. They were included to enable comparisons across teachers and class groups to test for any potential influence from primary school, be it classroom, primary school teacher and/or primary school achievement. It would be expected to see similar or weaker effects to the classroom analyses if primary school influences exist. If subject teachers have greater influence, larger effects would be anticipated here.

Maths and Geography Teachers

ANOVA results can be found for maths teachers in Tables 4.11 and 4.12

and for geography teachers in Tables 4.13 to 4.14. The results show for most of the measures, no significant effect of maths or geography teacher following multiple testing correction of $p \leq .001$ ($p = .05/41$). The results presented below show that measures that reached significance were mostly consistent with those that showed a significant effect of classroom, albeit with reduced effect sizes. Two exceptions just below the threshold were, maths homework behaviour ($p = .008$); and classroom chaos in both domains ($p = .009$). Levene's tests revealed equal variance for all measures except maths performance, number line task, geography primary school achievement, geography classroom environment, geography student-teacher relations, perceptions of academic and socioeconomic status and self-perceptions of family SES, occupation by geography teacher (see Appendix 4, Table 4.6 for maths teacher groups and Table 4.7 for geography teacher groups).

Maths primary school achievement. Students' end of year maths grade at primary school showed a modest effect of teacher, $F(5,201) = 4.634$, $p = .001$, $\eta_p^2 = .103$, with the highest mean score for TM5 and the lowest for TM4. Pairwise comparisons showed that students studying maths with teacher TM5 had significantly higher primary school maths achievement than students of TM4 ($p = .001$) following multiple testing correction of $p \leq .001$ ($p = .05/41$). Levene's test revealed equal variances were assumed for these analyses ($p = .41$).

Maths performance. A moderate effect of teacher was revealed, $F(5,223) = 11.697$, $p < .001$, $\eta_p^2 = .208$, with the highest mean score for TM5 and the lowest for TM6. Pairwise comparisons showed that students studying maths with teacher TM6 performed significantly lower than students of TM1 ($p = .001$), TM2 ($p < .001$), and TM5 ($p < .001$), following multiple testing correction of $p \leq .001$ ($p = .05/41$). However, Levene's test revealed unequal variances for these analyses ($p = .01$), with the smallest variance shown for teacher TM2 (0.38) and the largest for teacher TM3 (1.37).

Maths classroom environment. A modest effect of teacher was found, $F(5,217) = 4.700$, $p < .001$, $\eta_p^2 = .098$, with the highest mean score shown for TM5 and the lowest shown for TM1. This was the only significant pairwise comparison following multiple testing correction of $p \leq .001$ ($p = .05/41$), and revealed that students studying maths with teacher TM5 rated their classroom environment significantly higher than students of TM1 ($p < .001$). Levene's test revealed equal variances were assumed for these analyses ($p = .15$).

Maths student-teacher relations. A modest effect of teacher was observed, $F(5,217) = 5.468$, $p < .001$, $\eta_p^2 = .112$, with TM5 showing the highest mean score and TM1 showing the lowest. Pairwise comparisons demonstrated that students studying maths with teacher TM5 rated student-teacher relations significantly higher than students studying with TM1 ($p < .001$), and TM6 ($p < .001$), following multiple testing correction of $p \leq .001$ ($p = .05/41$). Levene's test revealed equal variances were assumed for these analyses ($p = .10$).

Perceptions of family SES – occupation by maths teacher. A modest effect of maths teacher was observed, $F(5,199) = 4.405$, $p < .001$, $\eta_p^2 = .100$, with TM4 showing the highest mean score and TM1 showing the lowest. No pairwise comparisons reached significance following multiple testing correction of $p \leq .001$ ($p = .05/41$). Levene's test revealed equal variances were assumed for these analyses ($p = .09$).

Table 4.11. Maths teacher groups: Means, standard deviation (SD) and N for maths classroom variables with ANOVA results by teacher group

Construct	TM1	TM2	TM3	TM4	TM5	TM6	*p*	η_p^2
Maths Primary school achievement	0.29 (0.96) n=16	-0.25 (0.96) n=19	-0.30 (1.04) n=30	-0.63 (0.79) n=18	0.48 (0.87) n=28	0.14 (0.94) n=96	**.001**	**.103**
Maths performance	0.66 (0.83) n=18	0.57 (0.62) n=25	0.06 (1.17) n=32	-0.34 (1.03) n=18	0.69 (0.80) n=28	-0.38 (0.87) n=108	**.000**	**.208**
Number line	-0.90 (1.19) n=18	-0.10 (0.64) n=25	0.08 (1.00) n=31	0.24 (0.71) n=18	-0.42 (1.17) n=28	0.03 (0.77) n=106	.234	.033
Maths self-perceived ability	-0.22 (0.99) n=16	-0.12 (1.05) n=22	0.06 (0.99) n=32	-0.01 (1.11) n=17	0.18 (0.88) n=28	0.03 (0.95) n=104	.809	.011
Maths enjoyment	0.15 (1.02) n=15	-0.07 (0.86) n=22	0.29 (0.80) n=31	0.22 (0.80) n=17	0.06 (0.89) n=28	0.00 (0.90) n=97	.593	.018
Maths classroom environment	-0.63 (0.80) n=17	-0.28 (0.86) n=22	0.03 (0.98) n=32	0.20 (0.86) n=17	0.61 (0.60) n=28	0.02 (0.98) n=107	**.000**	**.098**
Maths classroom student-teacher relations	-0.60 (0.94) n=17	-0.20 (0.87) n=22	-0.02 (1.01) n=32	0.23 (0.79) n=17	0.72 (0.58) n=28	-0.02 (0.96) n=107	**.000**	**.112**
Maths classroom peer competition	-0.04 (0.95) n=17	-0.24 (1.03) n=22	0.00 (0.93) n=31	0.03 (0.81) n=18	-0.11 (1.15) n=28	0.12 (0.94) n=107	.664	.015

Significant results in bold following a Bonferroni multiple testing correction of $p \leq .001$ where $p = .05$ divided by the number of measures (k=41) across maths and geography at time 1. Classes learning: English & Chinese = TM1; TM2; English = TM3; TM4; English and Spanish= TM5; TM6.

Table 4.11. Continued. Maths teacher groups: Means, standard deviation (SD) and N for maths classroom variables with ANOVA results by teacher group

Construct	TM1	TM2	TM3	TM4	TM5	TM6	*p*	η_p^2
Maths classroom chaos	0.27 (0.71) n=18	-0.05 (1.02) n=25	0.03 (0.98) n=32	0.02 (0.96) n=18	0.58 (0.78) n=28	-0.20 (1.05) n=106	.009	.066
Maths homework behaviour	-0.17 (0.85) n=18	0.39 (1.01) n=25	-0.04 (0.97) n=32	0.51 (1.01) n=18	-0.51 (1.04) n=28	-0.08 (1.06) n=106	.017	.061
Maths homework feedback	0.22 (0.87) n=18	-0.19 (0.88) n=25	-0.17 (1.02) n=32	-0.28 (1.35) n=18	0.14 (0.76) n=28	0.07 (1.02) n=106	.008	.067
Maths homework total scale	0.27 (0.65) n=18	-0.35 (0.99) n=25	-0.09 (0.90) n=32	-0.49 (1.27) n=18	0.30 (0.87) n=28	0.11 (0.94) n=105	.017	.060
Maths environment	-0.27 (1.15) n=17	-0.23 (1.04) n=23	0.20 (1.07) n=30	-0.16 (1.05) n=17	0.58 (0.69) n=28	-0.09 (0.96) n=104	.012	.066
Maths usefulness	0.13 (0.68) n=18	0.11 (0.93) n=24	-0.11 (0.78) n=31	0.07 (1.06) n=17	-0.41 (0.74) n=28	-0.02 (0.99) n=102	.295	.028
Maths anxiety	0.07 (0.95) n=17	-0.14 (0.99) n=24	-0.22 (0.95) n=32	-0.61 (0.78) n=17	0.00 (1.06) n=28	0.16 (0.96) n=102	.038	.053

Significant results in bold following a Bonferroni multiple testing correction of $p \leq .001$ where $p = .05$ divided by the number of measures (k=41) across maths and geography at time 1. Classes learning: English & Chinese = TM1; TM2; English = TM3; TM4; English and Spanish= TM5; TM6.

Table 4.12. Maths teacher groups: Means, standard deviation (SD) and N for perceptions of intelligence, and academic and socio-economic status variables with ANOVA results by teacher group

Construct	TM1	TM2	TM3	TM4	TM5	TM6	*p*	η_p^2
Theories of intelligence	0.07 (0.87) n=17	0.01 (1.06) n=25	0.03 (1.07) n=31	-0.24 (0.95) n=17	0.19 (0.90) n=28	-0.04 (1.03) n=102	.818	.010
Perceptions of academic and socio-economic status mean score	-0.40 (0.95) n=16	0.01 (0.98) n=24	0.11 (1.00) n=32	0.40 (1.09) n=17	0.03 (0.68) n=28	-0.05 (1.06) n=100	.301	.028
Perceptions Of School Respect	-0.09 (1.09) n=15	0.03 (1.22) n=23	-0.03 (0.93) n=30	0.04 (0.99) n=15	0.17 (0.93) n=26	-0.04 (0.98) n=95	.958	.005
Perceptions Of School grades	0.03 (0.87) n=16	-0.14 (1.03) n=23	0.05 (0.90) n=29	0.18 (1.07) n=17	-0.04 (1.02) n=26	-0.01 (1.04) n=98	.954	.005
Perceptions of family occupation	-0.76 (1.03) n=16	0.05 (0.82) n=24	-0.05 (0.95) n=31	0.76 (1.03) n=17	0.18 (0.66) n=25	-0.05 (1.03) n=92	**.001**	**.100**
Perceptions of family education	-0.49 (1.22) n=15	0.15 (0.92) n=22	0.28 (0.84) n=32	0.03 (0.93) n=16	-0.09 (0.91) n=26	-0.03 (1.05) n=94	.234	.033

Significant results in bold following a Bonferroni multiple testing correction of $p \leq .001$ where $p = .05$ divided by the number of measures (k=41) across maths and geography at time 1. Classes learning: English & Chinese = TM1; TM2; English = TM3; TM4; English and Spanish= TM5; TM6.

Table 4.13. Geography teacher groups: Means, standard deviation (SD) and N for geography classroom variables with ANOVA results by teacher group

Construct	TG1	TG2	TG3	TG4	TG5	*p*	η_p^2
Geography primary school achievement	0.00 (0.99) n=35	0.05 (1.01) n=69	0.67 (0.80) n=28	-0.31 (0.84) n=31	-0.18 (1.07) n=45	**.001**	**.083**
Geography performance	0.06 (0.84) n=43	0.06 (0.77) n=80	0.58 (0.68) n=28	-0.06 (0.88) n=32	0.00 (0.78) n=40	.016	.054
Geography self-perceived ability	0.07 (0.87) n=42	-0.13 (0.98) n=69	0.18 (1.03) n=26	0.05 (0.72) n=30	0.08 (1.16) n=44	.614	.013
Geography enjoyment	0.02 (0.99) n=41	-0.14 (1.00) n=71	0.15 (1.07) n=27	0.13 (0.89) n=30	0.03 (1.05) n=42	.618	.013
Geography classroom environment	-0.22 (1.00) n=40	-0.06 (1.04) n=73	0.82 (0.49) n=28	-0.10 (0.87) n=30	-0.15 (1.03) n=43	**.000**	**.104**
Geography classroom student-teacher relations	-0.21 (0.91) n=40	-0.02 (0.98) n=73	0.91 (0.56) n=28	-0.21 (0.97) n=30	-0.21 (1.06) n=43	**.000**	**.132**
Geography classroom peer competition	-0.20 (1.07) n=40	0.10 (1.02) n=73	-0.03 (0.88) n=28	0.06 (0.89) n=30	0.00 (1.04) n=43	.648	.012

Significant results in bold following a Bonferroni multiple testing correction of $p \leq .001$ where $p = .05$ divided by the number of measures (k=41) across maths and geography at time 1. Classes learning: English & Chinese = TG1; English = TG4; TG5; English and Spanish= TG2; TG3; TG5.

Table 4.13. Continued. Geography teacher groups: Means, standard deviation (SD) and N for geography classroom variables with ANOVA results by teacher group

Construct	TG1	TG2	TG3	TG4	TG5	*p*	η_p^2
Geography classroom chaos	0.08 (0.98) n=42	-0.19 (0.93) n=73	0.59 (0.72) n=28	-0.04 (1.13) n=31	-0.12 (1.07) n=44	.009	.061
Geography homework behaviour	0.20 (0.95) n=42	0.02 (1.10) n=72	-0.27 (0.88) n=28	-0.09 (0.97) n=31	-0.03 (1.02) n=43	.422	.018
Geography homework feedback	0.11 (0.95) n=42	0.05 (0.97) n=72	0.39 (0.78) n=28	-0.11 (0.89) n=31	-0.36 (1.18) n=43	.025	.051
Geography homework total scale	0.00 (0.98) n=42	0.00 (1.00) n=72	0.44 (0.79) n=28	-0.03 (0.89) n=31	-0.26 (1.14) n=43	.078	.039
Geography environment	0.14 (1.00) n=39	-0.05 (0.88) n=69	0.77 (0.70) n=28	-0.56 (1.04) n=28	-0.19 (1.02) n=43	**.000**	**.137**
Geography usefulness	0.37 (0.83) n=41	0.01 (0.91) n=73	0.09 (1.39) n=28	-0.28 (0.96) n=30	-0.25 (0.93) n=41	.028	.051
Geography anxiety	0.01 (1.00) n=38	0.05 (0.94) n=70	0.13 (0.81) n=28	-0.46 (0.73) n=30	-0.04 (1.00) n=41	.099	.038

Significant results in bold following a Bonferroni multiple testing correction of $p \leq .001$ where $p = .05$ divided by the number of measures (k=41) across maths and geography at time 1. Classes learning: English & Chinese = TG1; English = TG4; TG5; English and Spanish= TG2; TG3; TG5.

Table 4.14. Geography teacher groups: Means, standard deviation (SD) and N for perceptions of intelligence and academic and socio-economic status variables with ANOVA results by teacher group

Construct	TG1	TG2	TG3	TG4	TG5	*p*	η_p^2
Theories of intelligence	0.03 (0.98) n=42	0.01 (0.99) n=75	0.19 (0.90) n=28	0.03 (1.07) n=31	-0.19 (1.05) n=44	.596	.013
Perceptions of academic and socio-economic status mean score	-0.16 (0.98) n=40	-0.02 (1.00) n=73	0.03 (0.68) n=28	0.11 (1.00) n=32	0.07 (1.19) n=44	.801	.008
Perceptions Of School Respect	-0.01 (1.16) n=38	0.00 (0.95) n=70	0.17 (0.93) n=26	-0.03 (0.93) n=30	-0.07 (1.05) n=40	.919	.005
Perceptions Of School Grades	-0.07 (0.96) n=39	0.07 (0.96) n=72	-0.04 (1.02) n=26	0.05 (0.90) n=29	-0.07 (1.17) n=43	.926	.004
Perceptions of family occupation	-0.27 (0.98) n=40	-0.12 (1.06) n=68	0.18 (0.66) n=25	-0.05 (0.95) n=31	0.40 (1.02) n=41	.020	.056
Perceptions of family education	-0.11 (1.08) n=37	-0.05 (0.99) n=70	-0.09 (0.91) n=26	0.28 (0.84) n=32	0.03 (1.10) n=40	.515	.016

Significant results in bold following a Bonferroni multiple testing correction of $p \leq .001$ where $p = .05$ divided by the number of measures (k=41) across maths and geography at time 1. Classes learning: English & Chinese = TG1; English = TG4; TG5; English and Spanish= TG2; TG3; TG5.

Geography primary school achievement. A modest effect of teacher was found $F(4,203) = 4.586$, $p = .001$, $\eta_p^2 = .083$, with the highest mean score revealed for TG3 and the lowest for TG4. Pairwise comparisons demonstrated that students studying geography with teacher TG3 had significantly higher primary school achievement than students of TG4 ($p < .001$), and was the only significant comparison following multiple testing correction of $p \leq .001$ ($p = .05/41$). However, Levene's test revealed unequal variances for these analyses ($p = .05$), with the smallest variance shown for teacher TG3 (0.64) and the largest for teacher TG5 (1.14).

Geography classroom environment. A modest effect of teacher was observed, $F(4,209) = 6.086$, $p = .001$, $\eta_p^2 = .104$, with the highest mean score shown for TG3 and the lowest shown for TG1. Pairwise comparisons showed students studying geography with teacher TG3 rated their classroom environment significantly higher than students of TG1, TG2, TG4 and TG5 ($p < .001$), following multiple testing correction of $p \leq .001$ ($p = .05/41$). However, Levene's test revealed unequal variances for these analyses ($p = .009$), with the smallest variance shown for teacher TG3 (0.24) and the largest for teacher TG2 (1.08).

Geography student-teacher relations. The pattern of results, highly similar to classroom environment, revealed a modest effect of teacher, $F(4,209) = 7.943$, $p < .001$, $\eta_p^2 = .132$, with the highest mean score shown for TG3 and the lowest shown for TG1. Pairwise comparisons revealed that students studying geography with teacher TG3 rated student-teacher relations significantly higher than students taught by the other four teachers ($p < .001$), following multiple testing correction of $p \leq .001$ ($p = .05/41$). However, Levene's test revealed unequal variances for these analyses ($p = .012$), with the smallest variance shown for teacher TG3 (0.31) and the largest for teacher TG2 (0.96).

Geography environment. A modest effect of teacher was observed, $F(4,202) = 7.996$, $p < .001$, $\eta_p^2 = .137$, with the highest mean score found for TG3 and the lowest found for TG4. Pairwise comparisons showed that students studying geography with teacher TG3 rated their geography learning environment significantly higher than students of TG2, TG4, and TG5 ($p < .001$) following multiple testing correction of $p \leq .001$ ($p = .05/41$). Levene's test revealed equal variances were assumed for these analyses ($p = .54$).

Teacher Group Ranking by Mean Score

As with the classrooms, measures showing a significant effect of teacher group were also ranked by their mean scores (highest to lowest) to establish correspondence of rank across measures and across domains for these groups. If the influence of subject teacher is strong, a large amount of correspondence of rank would be expected for all teacher groups across the classroom measures within each domain. If the classroom influence is stronger, more variation in ranking for teachers with more classes might be expected.

Maths teachers. Table 4.15 shows slightly more consistency of rank for teacher groups across the measures (maths primary school achievement, maths performance, classroom environment, student-teacher relations, perceptions of family occupation, and classroom chaos - just below significance) compared to classrooms. Complete correspondence was observed between classroom environment and student-teacher relations. Of note is teacher TM5, who is ranked in first place across almost all measures. Although teachers covering more classes show less correspondence of rank across measures (e.g. TM6 teaches 4 classes), teacher TM4, who teaches one class also ranks inconsistently across the measures. Teacher TM5, who also teaches just one class of students, shows the most consistency across measures. There appears to be little relation then, between number of classes taught and amount of variation in ranking. However, recall that pairwise comparisons for primary school maths achievement showed that teacher TM5 inherited a class with the highest primary maths achievement and teacher TM4 inherited a class with the lowest primary maths grades. This could mean that ranking positions for these two teachers are partly due to prior achievement rather than any strong effect of maths teacher. While this explanation holds for performance and achievement, when rank is considered across other measures for teacher TM4, their students' have rated them highly for classroom environment and student-teacher relations.

Geography teachers. Similarly to maths, Table 4.16 also shows slightly more consistency of rank for teacher groups across geography measures (geography primary school achievement, classroom environment, student-teacher relations, geography environment, and classroom chaos – just below significance) compared to classrooms. Most consistent is teacher TG3 in first place across all measures. Teacher TG5 is also consistent for four out of five measures in fourth place. Teacher TG2 is consistently in second place across three measures. Correspondence is not complete between classroom environment and student-teacher relations as consistency in rank is only seen for three out of five teachers across the two subscales. As shown for maths teachers, there also appears to be no relation between number of classes taught and amount of variation in ranking.

Table 4.15. Maths teacher groups ranked by means (highest = 1 to lowest = 6) for measures demonstrating a significant effect of maths teacher

Construct	TM1	TM2	TM3	TM4	TM5	TM6	*p*	η_p^2
Maths Primary school achievement	**2nd**	**4th**	**5th**	**6th**	**1st**	**3rd**	**.001**	**.103**
Maths performance	**2nd**	**3rd**	**4th**	**5th**	**1st**	**6th**	**.000**	**.208**
Maths classroom environment	**6th**	**5th**	**3rd**	**2nd**	**1st**	**4th**	**.000**	**.098**
Maths classroom student-teacher relations	**6th**	**5th**	**3rd**	**2nd**	**1st**	**4th**	**.000**	**.112**
Perceptions of family occupation	**6th**	**3rd**	**4th**	**1st**	**2nd**	**5th**	**.001**	**.100**
Maths classroom chaos*	**2nd**	**5th**	**3rd**	**4th**	**1st**	**6th**	.009	.066

Significant results in bold following a Bonferroni multiple testing correction of $p \leq .001$ where $p = .05$ divided by the number of measures (k=41) across maths and geography at time 1. Classes learning: English & Chinese = TM1; TM2; English = TM3; TM4; English and Spanish= TM5; TM6.
*Not significant but used to make comparison with other measures

Table 4.16. Geography teacher groups ranked by means (highest = 1 to lowest = 5) for measures demonstrating a significant effect of geography teacher

Construct	TG1	TG2	TG3	TG4	TG5	*p*	η_p^2
Geography primary school achievement	**3rd**	**2nd**	**1st**	**5th**	**4th**	**.001**	**.083**
Geography classroom environment	**5th**	**2nd**	**1st**	**3rd**	**4th**	**.000**	**.104**
Geography classroom student-teacher relations	**5th**	**2nd**	**1st**	**4th**	**3rd**	**.000**	**.132**
Geography classroom chaos*	**2nd**	**5th**	**1st**	**3rd**	**4th**	.009	.061
Geography environment	**2nd**	**3rd**	**1st**	**5th**	**4th**	**.000**	**.137**

Significant results in bold following a Bonferroni multiple testing correction of $p \leq .001$ where $p = .05$ divided by the number of measures (k=41)
Across maths and geography at time 1. Classes learning: English & Chinese = TG1; English = TG4; TG5; English and Spanish= TG2; TG3; TG5
*Not significant but used to make comparison with other measures.

Relationships Across Domains (Mathematics And Geography)

Combined with the results from the ANOVAs, the consistency of class and teacher rank across the measures and across domains show some influence of teacher/classroom on these measures. Which has the most impact is unclear. If the subject teacher has greater influence, then a weak correlation would be shown between corresponding measures across the two domains (e.g. maths and geography performance). If peer group or primary school teacher have greater influence then a strong correlation would be observed between corresponding measures across the domains. To establish any underlying influence, bivariate correlations were estimated between the following corresponding measures that revealed a significant effect of maths and geography classroom or teacher group at time 1: maths and geography primary school achievement, maths and geography performance, classroom environment, student-teacher relations, classroom chaos, and maths/geography environment.

Table 4.17 shows moderate to strong correlations for the maths and geography pairs at each wave ranging from $r = .321$ to $r = .634$; the highest was shown for primary school achievement. The strength of the correlations between the pairs suggests negligible influence of subject teacher on the measures. The results imply a stronger effect of primary school teacher, although peers and prior achievement may also be confounding factors.

Table 4.17. Bivariate correlations (N) between maths and geography measures that demonstrated a significant effect of classroom and teacher at time 1

	Maths primary school achievement	Maths PVT	Maths classroom environment	Maths student-teacher relations	Maths classroom chaos	Maths environment
Maths primary school achievement	1					
	(219)					
Maths PVT	.310**	1				
	(207)	(229)				
Maths classroom environment	.160*	.117	1			
	(202)	(223)	(223)			
Maths student-teacher relations	.155*	.155*	.943**	1		
	(202)	(223)	(223)	(223)		
Maths classroom chaos	.035	.182**	.120	.197**	1	
	(205)	(227)	(221)	(221)	(227)	
Maths environment	.042	.087	.380**	.329**	.082	1
	(199)	(219)	(217)	(217)	(219)	(219)
Geography primary school achievement	**.634****	.230**	.124	.133	.144*	.168*
	(219)	(208)	(203)	(203)	(206)	(199)
Geography PVT	.245**	**.402****	.135*	.185**	.233**	.164*
	(201)	**(223)**	(217)	(217)	(221)	(213)
Geography classroom environment	.064	.141*	**.609****	.594**	.212**	.310**
	(193)	(214)	**(212)**	(212)	(214)	(211)
Geography student-teacher relations	.101	.122	.559**	**.584****	.202**	.220**
	(193)	(214)	(212)	**(212)**	(214)	(211)
Geography classroom chaos	.267**	.299**	.180**	.254**	**.631****	.093
	(197)	(218)	(213)	(213)	**(218)**	(212)
Geography environment	.082	.139*	.253**	.267**	.115	**.321****
	(185)	(207)	(204)	(204)	(207)	**(202)**

**Correlation is significant at the 0.01 level (2-tailed). *Correlation is significant at the 0.05 level (2-tailed). **Bold** indicates corresponding measure in each domain.

Comparison With Primary School

As the previous analyses suggested little influence from the subject teacher, potential influence from the primary school teacher and/or class was evaluated. Classroom and teacher group rankings of maths and geography primary school achievement were compared with those of the study measures and compared with each other. The relevant results are presented in Tables 4.8 to 4.10 for classrooms and Tables 4.15 and 4.16 for teacher groups.

Maths classroom. The results for maths class show some effect of primary school teacher/class when taking correspondence of rank for all the measures into account (see Table 4.8 and 4.9). Consistency of rank is shown slightly more frequently between primary school achievement and some of the study measures for most classes, with similarity of rank shown for the remaining classes. For example, for school achievement and student-teacher relations, C1e is consistently in 7^{th} place, C4se is in 5^{th} place, and C5se is in 6^{th} place. Correspondence is also seen between primary school achievement and classroom chaos for C6se and C7se in 2^{nd} and 3^{rd} places respectively (a high score indicates low chaos). Class C11ce in school 2 also shows consistency of rank across these two measures, in 3rd place.

Geography classroom. A similar pattern is shown for class ranking between geography primary school achievement and the study measures (see Table 4.10). Class C1e is consistently in 7^{th} place for school achievement, classroom environment and classroom chaos (a low score indicates high chaos). Class C6se is consistently in 1^{st} place across school achievement, classroom environment, student-teacher relations and geography environment. Class C5se is in 5^{th} place for school achievement and geography environment. Class C7se is in 6^{th} place across school achievement and student–teacher relations.

Maths and geography achievement. Considering all subjects at primary school level are taught by the same class teacher, we might expect to see substantial correspondence of class rank across maths and geography primary school achievement that goes beyond the well established correlation in performance across different domains, irrespective of teacher. For example, reading, mathematics and science have been shown to correlate highly (approximately .7) when taught by different teachers (Krapohl et al., 2014). As these correlations are less than unity, it implies other factors contribute towards variation in achievement across these subjects, factors that may include teacher/classroom effects. The high correlation across subjects has been shown to be largely due to substantial genetic overlap across the different domains. For example, the genetic correlation of 0.74 has been observed between reading and mathematics (Kovas, Harlaar, Petrill & Plomin, 2005), inline with the ‘generalist genes’ hypothesis, whereby the same genes contribute towards different traits (Kovas & Plomin, 2006). The results across maths and geography primary school achievement show some variation, with complete correspondence of rank for just three classes: C1e, C2e and C3e. Three other classes rank very closely: C5se, C6se and C8se; but the remaining two are a few ranks apart: C4se and C7se.

Maths and geography teacher groups. When we consider the teacher group rankings in Table 4.15, we can also see some variation between maths primary school achievement and the maths measures that showed a significant effect of teacher group. Correspondence across primary school achievement and the study measures is revealed for two out of five teacher groups, but as seen with the classroom ranks, some inconsistency is observed. For geography teacher groups (Table 4.16), slightly more correspondence of rank is observed between geography primary school achievement and the geography measures, compared to maths as three out of five groups are consistent.

When considering the primary school achievement rankings for classroom and teacher groups, the findings across both domains suggest some effect of primary school teacher and/or class between the groups as there is slightly more consistency between rankings for primary school achievement and rankings of the study measures, compared to the amount of consistency just within the study measures. The pattern, however, remains comparable with the study measures as correspondence is mainly seen for certain classes and for specific measures. The lack of complete correspondence between maths and geography primary school achievement across the class groups indicates that while there may be some influence of primary school teacher/class, other factors, perhaps pertaining to the subjects may have a greater influence. The slight variation between the two subjects may be due in part to variation in ability.

Differences In Primary School Achievement By Linguistic Specialism

To establish whether differences between primary school subjects assessed here are influenced by variation in ability, further analyses were conducted in relation to linguistic specialism. For example, the differences may reflect some ‘informal selection’ where parents enroll children in specialist language schools based on their child’s or their own characteristics. The differences may even reflect the actual effect of language (e.g. learning Chinese). The differences may alternatively, be due to an effect of school. ANOVA were conducted separately on maths and geography primary school achievement by language specialism (3 groups: English; English and Spanish;

English and Chinese), the results are presented in Table 4.18. For these analyses a Bonferroni multiple testing correction was set of $p \leq .025$ where p = .05 divided by the number of measures (k=2) across maths and geography at time 1. Levene's tests revealed unequal variance for these analyses (see Appendix 4, Table 4.8).

Table 4.18. Means, standard deviation (SD) and N for primary school achievement by language specialism with ANOVA results for language specialism

Construct	E	E&S	E&C	*p*	η_p^2
Maths Primary school achievement	-0.42 (0.96) n=48	0.22 (0.93) n=124	0.00 (0.99) n=35	**.001**	**.072**
Geography Primary school achievement	-0.55 (0.86) n=49	0.25 (0.98) n=124	0.00 (0.99) n=35	**.000**	**.108**

Significant results in bold following a Bonferroni multiple testing correction of $p \leq .025$ where p= .05 divided by the number of measures (k=2) across maths and geography at time 1.
Classes learning: E =English; E&S = English and Spanish; E&C = English and Chinese.

Maths primary school achievement. A modest effect of linguistic specialism was observed $F(2,204) = 7.857$, $p = .001$, $\eta_p^2 = .072$, with the highest mean score revealed for the group learning English and Spanish and the lowest for the group learning English. Pairwise comparisons demonstrated that students studying English and Spanish had significantly higher primary school maths grades than students learning just English ($p = .001$), following multiple testing correction of $p \leq .025$ ($p = .05/2$). No difference was revealed between the English and Chinese linguistic group and the other two groups. However, Levene's tests revealed unequal variance for these analyses ($p = .04$), with the smallest variance shown for the English and Spanish group (0.86) and the largest for the English and Chinese group (0.98).

Geography primary school achievement. Similarly to maths, a modest effect of linguistic specialism was observed $F(2,205) = 12.423$, $p < .001$, η_p^2=

.108, with the highest mean score again revealed for the group learning English and Spanish and the lowest for the group learning English. Pairwise comparisons demonstrated that students studying English and Spanish had significantly higher geography primary school grades than students learning English ($p < .001$), following multiple testing correction of $p \leq .025$ ($p = .05/2$). Again, no difference was revealed between the English and Chinese linguistic group and the other two groups. However, Levene's tests revealed unequal variance for these analyses ($p = .013$), with the smallest variance shown for the English learning group (0.74) and the largest for the English and Chinese group (0.98).

As the significant difference between linguistic specialisms is shown only between students learning English and students learning English and Spanish, but not between students learning English and students learning English and Chinese, it suggests that the difference is not necessarily due to learning two languages compared to one. Additionally, as the difference was observed between two linguistic groups within the same school, no effect of school was revealed.

Summary

To summarise, the majority of measures across maths and geography classrooms, in school 1 and school 2, showed no significant effect of classroom or teacher. Some measures that demonstrated a significant effect of classroom and teacher were significant for both mathematics and geography contexts. These effects were for achievement, performance, classroom environment, classroom atmosphere and student-teacher relations. No teacher or classroom effects were found for motivation, homework behaviour/feedback and subject anxiety. The effect sizes observed for teacher effects (8.3% to 20.8%) were smaller compared with classroom effects (12.8% to 34.7%). A modest significant effect of linguistic specialism was found for primary school achievement with students learning English and Spanish combined demonstrating the highest mean score. Surprisingly, no differences were shown for the group learning English and Chinese.

The ranking showed variability across measures for most classrooms. However, specific classes showed some consistency across measures and across maths and geography. Slightly more consistency in ranking was exhibited for teacher groups within maths and geography measures; however, complete correspondence was not found. Slightly more correspondence was shown with primary school subjects suggesting that any teacher/classroom influences stem from primary school.

'Teacher/classroom effects' presented in this study refer to statistical significance of the comparison of the groups by current subject teacher. This, however, does not mean actual effect, as the results may be confounded by other factors, such as prior class achievement.

Discussion

The main aim of the present study was to investigate whether being in the same classroom with the same peers during primary and secondary education would lead to a significant effect of teacher/classroom on measures of school achievement, performance, classroom environment, motivation and subject anxiety. No significant effect of classroom was found for the majority of constructs. Only ten measures, from a total of 35 across maths and geography classrooms, showed significant differences and these were mainly for school 1 (see Tables 4.2 to 4.6). These measures were similar for the two domains, and relate to school achievement, classroom environment, classroom atmosphere and student-teacher relations. There were just two exceptions; maths performance and geography environment were significantly different across classrooms for each domain only. Effect sizes were moderate, ranging from 12.8% to 34.7%.

Because one teacher at secondary education teaches several classes, this enabled the investigation to disentangle teacher and classroom effects by regrouping the students by each teacher. The findings were highly similar to the classroom results with significant effects of teacher found for the same measures of achievement and teacher/classroom environment across maths and geography. It can be seen, however, that the effect sizes for teacher groups ranging from 8.3% to 20.8% (Tables 4.11 to 4.14), are slightly smaller compared to those shown for classroom effects, which ranged from 12.8% to 34.7% (see Tables 4.2 to 4.6). The smaller effect of subject teacher on primary school achievement would be expected considering this subject was taught by the primary school teacher and not the current subject teacher tested here. The smallest reduction in effect size was observed for maths student-teacher relations. This is also anticipated considering that students were rating their maths teacher, and therefore one would expect the teacher to contribute substantially towards the effect size on this measure. However, for geography the situation is somewhat different as there was a larger reduction in teacher effect size (13.2%) compared to classroom (18.9%) suggesting less impact of the subject teacher here, comparatively. Overall, these findings suggest some influence of subject teacher but the impact of classroom, and other potential factors, is slightly larger.

As students may have been enrolled in programmes to learn additional languages on the basis of ability, linguistic specialism was investigated to separate any influence of implicit selection. The results showed a modest effect (7.2% to 10.8%) of linguistic specialism on primary school achievement (see Table 4.18). Students studying English and Spanish had significantly higher maths and geography primary school achievement than students studying just English. The difference was only shown between these two groups, no effect was shown between the English and Chinese group and students learning English. This suggests little or no impact of learning two languages compared to one. Unless the additional cognitive load associated with learning two such diverse languages prevented the English and Chinese learners from gaining significantly higher primary school achievement. These findings suggest that differences observed for the English and Spanish group are more likely to be driven by other factors relating to the teacher or class group rather than factors associated with their choice of linguistic specialism and/or language ability. In addition, as the effect was shown between the two linguistic groups within the same school, this suggests no effect of school.

With the effect of classroom and teacher being specific to performance and teacher/classroom environment, consistency of rank was explored to assess whether the influence of classroom and subject teacher was constant across all measures. A weak effect was observed overall as correspondence of rank across measures and within domains was shown only for certain classes and teacher groups (see Tables 4.8 to 4.10 for classrooms; and Tables 4.15

and 4.16 for teachers). There appeared to be slightly more consistency for teacher groups across measures and domains than shown for classrooms. This may suggest a stronger influence of teacher; however, this may be due to fewer numbers of groups to be ranked, compared to classrooms, allowing for fewer permutations than might be observed with a larger number of groups. Amount of variation in rank across measures was not specific to the number of classes taught, as some variation was shown for teachers with any number of class groups. If the effect of subject teacher was stronger, less variation would have been observed. This suggests classrooms and other factors are driving the differences, rather than subject teachers. The consistency of rank across the two domains shown for certain class and teacher groups also suggests the contribution of other factors.

Indeed, the moderate to strong correlations found between the measures also suggested a negligible effect of the current subject teacher (see Table 4.17). Strong correlations between maths and geography primary school achievement would be expected when these stem from the same classroom/teacher. Especially given the large correlations evidenced across maths, English and science taught by different teachers (e.g. Krapohl et al., 2014). However, the strong correlations shown across maths and geography classrooms for other constructs: classroom environment, student-teacher relations and classroom chaos, also indicates little influence from the subject teacher. This was further substantiated with slightly more consistency in rankings seen between primary school achievement and some of the study measures. However, these ranks only corresponded for specific classes and teacher groups and in many cases were not consistent across all measures.

Signifying that while current subject teacher effects are weak, primary school teacher/class effects are not overwhelmingly strong in their absence. Equally, complete correspondence of rank was not found between maths and geography primary school achievement, which is surprising considering all subjects at this level are taught by the same teacher. If an overriding effect of teacher was to be found, it would have been revealed here. The amount of correspondence observed across domains may reflect established correlations that suggest associations between subjects are strong beyond any effect of the teacher and/or class (Krapohl et al., 2014). The absence of complete correspondence across maths and geography primary school achievement may suggest some variability in teacher and/or student proficiency in relation to the two subjects. For example, given the number of subjects taught by a primary school teacher, it is reasonable to expect they may be more proficient in teaching some subjects compared to others. Likewise, differences between subjects may be due to variation in student ability across the two domains.

Of interest is one particular class group (C6se) that in addition to maintaining first place across maths and geography classrooms for the majority of measures, also ranked highly for teachers. This group also demonstrated more frequently, significantly higher mean scores compared to other classes/groups across pairwise comparisons. This class is taught by maths and geography teachers who teach no other classes in this year group (TM5 and TG3). Being the only class for this year group is unlikely to be a factor though. Teaching load is comparable across teachers, as they will also teach other years' classes across the school. When considering this classes' linguistic specialism, however, it might be suggested that learning English and Spanish is a factor in higher achievement especially as the class ranking lowest most frequently is learning just one language. Additionally, as a group, students learning both English and Spanish are doing significantly better. It is also feasible that parents have sent ‘stronger’ students to learn two languages. However, one might also have expected classes learning English and Chinese to be doing significantly better, considering the challenge of learning two new language systems that are entirely different to their own. The absence of effect for students learning English and Chinese, however, suggests that learning two languages per se is an unlikely driver of effects for this specific class. Instead, it might be that the high ability of this class is driving the significant effects observed for the group learning English and Spanish rather than the reverse. As Levene’s tests revealed unequal variances for many of the measures that showed a significant effect of classroom, teacher and linguistic group, it not only prevents complete confidence in interpreting the results, it also might be expected that a few brighter children are influencing the performance of this particular group. However, as class C6se and teacher TG3 was most often the class and teacher group with the smallest amount of variance it suggests greater similarity within this classroom in high ability and good student-teacher relations.

The nature of effects demonstrated by the study are interesting in that modest to moderate effects of teacher/classroom were shown for measures associated with classroom and teacher environment as opposed to self-perceived ability, subject enjoyment and maths or geography anxiety. It appears that being in a particular classroom with a specific teacher did not significantly influence variation in student motivation or subject anxiety. Remarkably, considering the classes are mixed ability, there is substantial evidence that some classes are doing significantly better and others are doing significantly worse. It is also apparent that classes doing better have lower levels of classroom chaos and higher levels of student-teacher relations. The converse is true for students at the lower end of the achievement scale. These results offer some support for findings where prior achievement moderated student-teacher relations and led to greater academic respect and acceptance among peers for high and low ability students (Hughes et al., 2014). It may be that student engagement is being enabled by greater teacher support in these specific groups as found in previous research (Wang & Eccles, 2012). Correspondingly, for students doing worse, the relationship with their teacher may be such that a good emotional climate is not sustained within their class group (e.g. Reyes et al., 2012). Where class groups are doing well, the increased mean scores may be the result of average ability students doing better when among higher achieving peers (e.g. Carmen & Zhang, 2012). However, in the classes doing less well, lower ability students may be less receptive to any influence from higher peer achievement (e.g. Carmen & Zhang, 2012). Equally, if peers can influence either positively or negatively, it may be this factor that is increasing or decreasing student outcomes (e.g. Haworth et al., 2013; Wang & Eccles, 2012). These findings provide some insight into the complex nature of teacher/class effects and how they are subject to several confounding factors.

Strengths and Limitations

This study is not without limitations. One issue is the time of data collection. The initial plan was to make the first data collection at the beginning of the autumn term before any influence from the new subject teachers and timetable took hold. Due to some practical issues, the data were collected at the beginning of the spring term. Small group sizes were also seen in some cases. A certain amount of attrition is expected but it is unclear whether participation is completely random for such classes or whether only selected ability students are taking part. One other factor is the number of measures used. This is both a strength and a limitation, as on the one hand it enabled the testing of multiple constructs within the classroom environment, but on the other hand, required the application of a stringent multiple testing correction across analyses: more constructs would have been significant if fewer measures were used. However, this study highlights the complexity of within-classroom factors rather than focusing on just one or two aspects.

Conclusion

In conclusion, these findings suggest that for some students, being in a particular class/teacher group has a moderate effect on measures of school achievement, performance, class environment and student-teacher relations. As the effect of teacher group is somewhat reduced compared to class group, this suggests other factors contribute. The moderate to strong correlations found between the measures across the two domains also indicates a negligible effect of subject teacher. It may be that being among the same peers with the same primary school teacher for four years of education has some influence beyond the modest effect of the current subject teacher. The level of correspondence in rank between the study measures and primary school achievement offers some support to the idea of the primary school teacher setting a class ethos that is unchangeable by the current subject teacher. Considering though that rankings are not completely consistent across measures and domains for all class and teacher groups, it suggests the involvement of additional influences. The contribution of variation in student ability and implicit selection processes cannot be discounted. These findings suggest a weak effect of subject teacher, confounded by multiple factors, many of which stem from primary school.

Chapter 5

Examining continuity of teacher and classroom influences from primary to secondary school

Abstract

Significant effects of classrooms and teacher groups found in Chapter 4 may erroneously be assumed to stem solely from teacher effects. However, they may also result from other factors, such as student characteristics, primary school factors or selection processes. To establish any influence from primary school, this study uses cross-sectional and longitudinal methods in two samples of 10 to 12 year old secondary school students, one from Russia and one from the UK. The results showed that significant effects of classroom and teacher groups found at time 1 for maths and geography educational outcomes continued at time 2 but weakened at time 3, especially for maths classrooms. Longitudinal analyses suggested a weak influence from primary school classrooms and teachers, that extended to time 3 for geography classrooms. The results suggest that multiple influences contribute towards classroom and teacher group variation. This should be taken into account by policymakers involved in teacher promotion and employment prospects.

Introduction

Following on from Chapter 4, this study aims to examine further the potential teacher/classroom effects on measures of school achievement, performance, classroom environment, motivation and subject anxiety at additional assessment waves across the academic year: at time 2 (April/May); and at time 3 (September, following the summer break). In Chapter 4, a significant effect of classroom/teacher at time 1 was observed for several measures. However, the effects may be confounded by other factors such as variation in student ability, peer influences, a classroom ethos set by the primary school teacher and/or implicit selection processes related to student enrollment into a more challenging language curricula. Assuming that the observed effects (average differences between classes in academic performance and other outcomes) are due to teacher influence has implications. In many countries, including the UK and Russia, the current policy is to base decisions concerning employment and promotion of teachers on ‘added value’ that teachers bring, beyond individual students’ characteristics.

Selection processes, whereby students are assigned to classrooms or schools on the basis of prior ability, have been shown to differentially influence students of different ability. Much of the literature suggests that they benefit higher ability students but are detrimental for students at the lower end of the ability spectrum (e.g. Burgess, Dickson, & Macmillan, 2014; Hattie, 2002; Kelly & Carbonara, 2012; Maaz, Trautwein, Ludtke & Baumert, 2008). Research has suggested however, that selection effects are also confounded by peer effects, whereby students perform better or worse depending on the ability level and work ethos of their fellow students (Guyon, Maurin, & McNally, 2012). It is a dilemma for policy makers as it seems that higher ability students benefit more from being among high achieving peers rather than among lower achieving ones; and students of average (Carmen & Zhang, 2012) and lower ability, do better in classes where they can mix with high ability peers (Ding & Lehrer, 2007).

Peer group and selection also have implications for students' mathematics self-concept. When tracked for just one or two specific subjects, low ability students were observed as having low maths confidence, and high ability students were observed as having high maths confidence (Chmielewski, Dumont, & Trautwein, 2013). However, when tracked for all subjects within a school, the situation was reversed and high ability students had lower maths confidence compared to their low ability counterparts. The study suggested that the daily regrouping of students for one or two specific subjects continually expanded the student's frame of reference for self-concept, so they would compare themself against the whole year group, therefore putting themselves at the extreme ends of a larger population. Whereas, being grouped with peers across all subjects, students continually maintained a smaller frame of reference to include only immediate peers (Chmielewski et al., 2013).

In the UK there is a high amount of selection and streaming for ability. In primary education (ages 5 to 10 years), state funded education is mixed ability, although there may be some setting and grouping by ability within classes. In secondary education (aged 11 to 18 years), the majority of schools are mixed ability although most still apply ability streaming for maths and English lessons. Other schools, select students on ability prior to enrollment when stringent testing takes place. Some boroughs implement testing at age 10 and students who pass have the opportunity to enroll in higher ability schools (grammar schools).

The UK system also gives a high degree of choice dependent on locality. Within a district there will be several state funded schools and each has a catchment area of a certain radius from which to take students. Catchment areas can vary between schools, depending on availability of places. Choice of school, which is not unique to the UK, leads to more implicit selection processes. Schools with a good reputation, attract parents in such a way that many will move house to the catchment area. This leads to the value of property increasing in the surrounding area (e.g. Figlio & Lucas, 2000; Gibbons, Machin & Silva, 2013). This in turn leads to school selection by affordability, leading to uneven distributions of parental SES within schools. These schools tend to become over-subscribed and in many cases, implement their own testing criteria to control applicant numbers. One such school known to the author, has good league table results and therefore is highly popular with parents. One selection criteria for this school is a smaller catchment area compared to other schools locally. Another criteria is the need for applicants to pass a test to obtain a place. The head-teacher sets the test date for 9 am on a Saturday morning, and anecdotally, it is said that the headmaster devised this as a preliminary selection process, in that only parents motivated enough to get their child to school early on a Saturday morning for the test need apply.

In Russia, there is no formal streaming (except for specialist music, art, maths, and other schools – for gifted students or special needs schools). Here however, there may still be some form of implicit selection of schools as they also have a degree of choice. Research in Toronto, Canada, has shown that when parents were given a choice of several high schools in their locality, 41 percent opted for a different school rather than send their child to the one they were initially assigned to (Leonard, 2011). Further, this decision was associated with student ability and perception of a potentially stronger peer group. Within the Russian sample in this study, where parents have chosen a specific linguistic pathway, there may also be some informal selection at play. Parents have enrolled their child into a school with a challenging programme to learn up to two second languages, which suggests a level of confidence in their child's capability.

By selecting a school, parents entrust the school with all aspects of their child's education, including who teaches them. At entry to primary school, UK students are randomly assigned to class groups and teachers. This is similar in Russia, although unofficially, parents may try to obtain a class place for their child with a certain teacher based on local reputation, especially if an older sibling has attended already. This indicates that certain teacher characteristics may influence parental choice.

Apart from perceived teacher performance in terms of student achievement, it is probable that other attributes of the teacher may influence parental choice. For example, having effective interpersonal skills, which indicates good emotional intelligence (Brackett, Mayer, & Warner, 2004) has been shown to lead to better student-teacher relations and a more favourable classroom climate (Maulana, Opdenakker, & Bosker, 2014; Reyes, Brackett, Rivers, White & Salovey, 2012).This in turn, has been shown to improve levels of student motivation, especially in the first year of secondary education (Maulana, Opdenakker, Stroet, & Bosker, 2013), when motivation levels have been shown to drop (Anderman, Maehr, & Midgley,1999). This first year of secondary education is suggested to be a particularly vulnerable stage of child development (e.g. Eccles, 1999), not helped by the transition from having one teacher during primary education to having multiple teachers across different subjects. Unsurprisingly, teacher-student relations have been shown to fluctuate across the academic year, in particular, the first year of secondary education (Maulana et al., 2013).

Teachers have a difficult task in engaging students at this stage. Their own self-efficacy in student engagement needs to be resilient (Tschannen-Moran, Hoy, & Hoy,1998). They may also need to be confident in their ability to employ a range of instructional strategies as well as utilize good classroom management skills in order to engage the students (Tschannen-Moran, & Hoy, 2007). These attributes are essential at the student and teacher level. In particular for students, as higher levels of maths performance have been observed in more orderly classrooms (Opdenakker, & Damme, 2001). In the case of teachers, self-efficacy has been shown to associate with job satisfaction (Caprara, Barbaranelli, Steca, & Malone, 2006), and teacher burnout (Skaalvik, & Skaalvik, 2007). Although research has shown no differences in self-efficacy in student engagement between teachers with high or low levels of experience, teachers with more experience revealed higher levels of self-efficacy in both classroom management and their use of instructional strategies (Tschannen-Moran, & Hoy, 2007).

The previous research presented here shows that students' educational outcomes can be influenced by several factors, for example, peer influences, implicit or explicit selection processes, classroom ethos set by the primary school teacher and/or prior achievement. In order to disentangle factors, potential effects need to be explored longitudinally, to see whether effects found at earlier assessment waves persist across the later waves. Additionally, they need to be investigated without any influence of prior achievement to control for any remaining influences from the primary school teacher/classroom, and primary school achievement. One way to further disentangle teacher influence, is to explore associations between teacher characteristics and measures which revealed a significant effect of classroom/teacher. By estimating the strength of relationships separately for current subject teacher and primary school teacher, it is possible to detect to some extent, the potential influence from the different teacher groups. Another way to untangle factors is to explore mediating influences between teacher characteristics and classroom measures/achievement.

The Current Study

The analyses reported in this chapter investigates continuity of effects (reported in Chapter 4) of maths and geography classrooms and teachers. Specifically, the study aims to investigate whether the significant effects found at time 1 (the first assessment wave in January) for achievement, performance, classroom environment and student-teacher relations persisted across the academic year at time 2 (April/May) and time 3 (September, following the summer break). The analyses are organized into five parts which are identified with the research questions and include separate discussions. Parts 5.1 and 5.2 assess whether any influences remained from primary school teacher/class and/or primary school achievement analyses were conducted with and without controlling for primary school achievement to compare any potential differences in effects between the sets of analyses. In part 5.3, analyses were repeated in the UK sample in order to compare patterns with a selective education system where formal streaming is implemented. In part 5.4 associations between academic outcomes and teacher characteristics were explored prior to mediation analyses. Part 5.5 presents mediation analyses conducted in the Russian sample to show any additional relationships between academic outcomes and teacher characteristics. The following research questions are addressed: 1) (Part 5.1) Do significant effects of classroom and teacher groups found at time 1 persist across time 2 and time 3? 2) In the case of patterns of significant effects persisting across time 2 and time 3, are the patterns of class rankings found at time 1 also maintained across subsequent waves? 3) (Part 5.2) Do potential patterns of significant effects and rankings persist when taking account of prior achievement? For example, do differences between classrooms in maths performance disappear once the differences in primary school achievement are accounted for? 4) Do potential significant effects and ranking patterns persist in the same way across maths and geography class and teacher groups at time 2 and time 3? 5) (Part 5.3) Are potential significant effects and ranking patterns found in the Russian sample similar to any potential effects found in the UK sample? For example, if significant effects of classroom are found for maths performance in the UK sample, is the strength of effect stronger or weaker than effects found in the Russian sample? Further, if effects are found for several measures, are ranking patterns more or less consistent than those found in the Russian sample? 6) (Part 5.4) In the Russian sample, do teacher characteristics associate with classroom environment measures and performance/achievement? For example, do primary school teacher characteristics associate more strongly or more weakly than current subject teacher characteristics? 7) (Part 5.5) In the Russian sample, do teacher characteristics mediate potential relationships between classroom environment measures and performance/achievement?

Methods

Participants

Students. All student participants were in the first year of their secondary education, with specific subject teachers for the first time. Both samples attended co-education schools (both sexes educated together).

Russia. Participants were 229 (102 males; 127 females) 10 to 12 year old students (mean age 142 months, range 127-156 months) from two urban mixed ability schools in St. Petersburg, Russia. The schools had specialized linguistic programmes that provided the students with the opportunity to learn up to two languages: English; English and Spanish; and English and Chinese. In one school (School 1), there were three classes of students who learned English and five classes of students who learned English/ Spanish. In the other school (School 2), there were three classes of students who learned English and Chinese. These 11 classes included all classes of this year group in the two schools. For further details, please see Chapter 4, p. 118.

UK. Participants were 163 (97 males; 66 females) 11 to 12 year old students (mean age 140 months, range 135-158 months) from one urban mixed ability school in London, UK. Although the school is mixed ability, students were streamed by ability for their maths classes. In year 7 (the first year of secondary education), there were four levels of ability (numbered from 1 (highest ability) to 4), with two classes at each level. Students with special educational needs were excluded from these analyses, many of whom were at level 4. At the beginning of year 8 (time 3 of the current study), students were redistributed by their school, so that more students were assigned to lower ability at level 4, which led to an extra class group at time 3. A total of six classes participated at time 1 and time 2, and eight classes at time 3. The same number of participants took part at each assessment, they were just distributed differently across the number of classes at time 3.

Teachers. Data were available from the Russian students' teachers (N = 17, all female), aged between 35 and 63 years (M = 49.93, *SD* = 7.87). Participating teachers included eight primary school teachers, six maths teachers and four geography teachers. One primary school teacher continued with the class into secondary education as their maths teacher.

Measures

A detailed description of the measures used in this study is provided in the methods section in Chapter 2, pages 61 to 72.

Procedure

A detailed description of the procedure used in this study is provided in Chapter 4, page 118.

Analyses

Analyses were conducted using data collected at the second and third assessments (time 2 and time 3) on variables corrected for age, with outliers (±3SD) removed. In addition, maths performance data at time 1 were also analysed.

Russia. Analyses of variance (ANOVA) were conducted within each of the two Russian schools to assess potential differences between classes in means for school achievement, performance, classroom environment, motivation, attitude towards subject, and subject anxiety at time 2 and time 3. A Bonferroni multiple testing correction was set of $p \leq .000$ where $p = .05$ divided by the number of measures (k=114) across both schools and maths and geography, time 2 and time 3, with one additional measure from time 1. The measures included: maths classroom measures = 14 x 4 (14 measures assessed at time 2 and time 3, separately within school 1 and school 2); geography classroom measures = 13 x 4 (13 measures assessed at time 2 and time 3, separately within school 1 and school 2); maths achievement =1 x 2 (1 measure available at time 2 only, assessed separately within school 1 and school 2); geography achievement = 1 x 2 (assessed at time 2 only, separately within school 1 and school 2); maths performance = 1 x 2 (assessed at time 1 only, separately within school 1 and school 2). Planned pairwise comparisons were conducted between classrooms applying a Dunnett's T3 multiple comparison correction as it maintains tight control of the Type 1 error rate while allowing for differences in variances and group size (Field, 2011).

To further investigate potential effects of the teacher/classroom, students' classes were regrouped to account for secondary school teachers teaching more than one class. In other words, all children taught, for example, by the same maths teacher, were grouped together when maths-related measures were analysed. ANOVAs by teacher group were conducted with both Russian schools combined, for maths and geography measures separately. For these analyses a Bonferroni multiple testing correction was set of $p \leq .001$, where $p = .05$ was divided by the number of measures (k=57) across the two sets of teachers (maths and geography) at time 1, time 2 and time 3. This translates as: maths classroom measures = 14 x 2; geography classroom measures = 13 x 2; maths achievement =1 x 2 (time 2 only); geography achievement = 1 x 2 (time 2 only); maths performance = 1 x 1 (time 1 only). Planned pairwise comparisons were also conducted between the teacher groups, using a Dunnett's T3 multiple comparison correction.

Following the observation of any significant differences between classes for any measures, the classes and teacher groups were ranked by means, highest to lowest, to assess any correspondence of class ranking across the significant measures. Although these analyses were conducted separately for maths and geography classroom measures, class ranking was compared across the two domains.

Further analyses were conducted to see whether any classroom/teacher effects were influenced by the primary school teacher, or other primary school factors, such as prior achievement, or implicit selection processes. All analyses (ANOVAs by teacher; by class and ranking) were repeated for maths and geography measures at time 2 and time 3 with all variables regressed on primary school achievement to control for prior achievement. In order to make a similar comparison with variables at time 1, maths performance at time 1 was also regressed on primary school achievement to control for prior achievement and the analyses were repeated.

Further analyses were conducted to investigate potential associations between academic outcomes and the following teacher characteristics: years of teaching experience; emotional ability; and self-efficacy in student engagement, classroom management and instructional strategies. Bivariate correlations were conducted between teacher characteristics and measures that demonstrated a significant effect of classroom at time 1, time 2 and time 3. These were conducted separately for teacher characteristics of primary teacher, maths teacher and geography teacher. This led to 111 sets of analyses that would require a Bonferroni multiple testing correction of $p \leq .0004$ where $p = .05$ divided by the number of measures (k=111). However, to maintain any potential meaningful associations that may otherwise be lost under such stringent correction, the decision was made not to apply the correction in these analyses. Instead, further replication of any significant findings will be necessary.

Mediation analyses were conducted to investigate whether significant associations between achievement/performance and classroom environment measures were mediated by teacher characteristics and/or primary school achievement. Nineteen simple mediating models were conducted using Mplus 7.4 (Muthen & Muthen, 2015). To prevent sample bias, bootstrapping with 5000 resampling was used (Geiser, 2013; Preacher & Hayes, 2008). Students' age was used as a covariate for the maths and geography classroom measures to control for potential age effects.

UK. To explore whether similar patterns of results would be found within a selective education system where formal streaming is implemented, analyses were conducted using data from the UK sample. These data were from assessments across the academic year that matched the Russian data collections (see procedure in Chapter 3). ANOVAs were conducted separately by maths and geography classroom without controlling for prior achievement, to assess potential differences in means for cognitive ability test, school achievement, performance, classroom environment, motivation, attitude towards subject, subject anxiety at time 1, time 2 and time 3 and perceptions of intelligence and socioeconomic status at time 1. A Bonferroni multiple testing correction was set of $p \leq .001$, where $p = .05$ was divided by the number of measures (k=98) across maths and geography at time 1, time 2 and time 3.This translates as: maths classroom measures = 14 x 3; geography classroom measures = 13 x 3; maths achievement =3 x 1 (2 for time 1;1 for time 2); perceptions of intelligence and socioeconomic status = 6 x 2 (time 1 only) cognitive ability = 1 x 2. Planned pairwise comparisons were conducted between classrooms applying a Dunnett's T3 multiple comparison correction as it maintains tight control of the Type 1 error rate while allowing for differences in variances and group size (Field, 2011).

Following the observation of any significant differences between the UK classes for any measures, ranking analyses by mean scores (highest to lowest) were also performed to assess any correspondence of class ranking across potential significant measures. As with the Russian analyses, ranking was conducted separately for maths and geography classroom measures and comparisons were also made across domains.

Results

5.1. Classroom And Teacher Differences, At Time 2 And Time 3 In The Russian Sample, Without Controlling For Prior Achievement

Means, standard deviations (SD) and N for all assessed variables by classroom and by teacher, are presented in Tables 5.1.1 to 5.1.6 for time 2 and in Tables 5.1.13 to 5.1.18 for time 3.

Differences between maths classes at time 2

School 1. ANOVA results by maths classroom at time 2,without controlling for prior achievement, are presented in Table 5.1.1 The results show for the majority of measures, no significant differences between maths classrooms following multiple testing correction of $p \leq .000$ ($p = .05/114$). Maths environment fell just below significance ($p = .001$). Levene's tests revealed that equal variances were assumed for most measures, except maths performance, classroom chaos, and homework feedback (see Appendix 5, Table 5.1.1). Levene's tests were not corrected for multiple testing, instead they were reported for each ANOVA separately with a significance level set at $p \leq .05$. For the following measures, significant average and variance differences between the classes were observed:

***Maths performance time 2*.** Modest significant differences between classrooms were found for maths performance, $F(7,177) = 4.158$, $p < .001$, $\eta_p^2 = .141$, with the highest mean score revealed for C6se and the lowest for C5se. Pairwise comparisons showed that this was the only significant difference, revealing that students in class C6se had significantly higher maths performance than students in class C5se ($p < .001$), following multiple testing correction of $p \leq .000$ ($p = .05/114$). However, Levene's test showed unequal variances for these analyses ($p = .041$). C6se had the least amount of variance, (0.47), C2e had the most variance (1.56).

***Number line time 2*.** A modest effect of classroom was found, $F(7,175) = 5.225$, $p < .001$, $\eta_p^2 = .173$, with the lowest mean score (optimum score) revealed for C4se and the highest for C8se. No significant pairwise comparisons were observed following multiple testing correction of $p \leq .000$ (p = .05/114). Levene's test showed equal variances were assumed for these analyses (p = .203).

***Maths classroom chaos time 2*.** Modest significant differences between classrooms were found for classroom chaos, $F(7,178) = 6.222$, $p < .001$, $\eta_p^2 = .197$, with the highest mean score (low chaos) revealed for C7se and the lowest (high chaos) for C4se. Pairwise comparisons revealed that C7se was significantly higher than C5se only ($p < .001$), following multiple testing correction of $p \leq .000$ (p = .05/114). This means that class C7se students' perceptions of chaos levels were significantly lower than students' perceptions of chaos levels in class C5se. However, Levene's test showed unequal variances for these analyses (p = .024), which likely explains the significant pairwise comparison falling between C7se and C5se instead of between C7se and C4se. Variances for C7se and C5se were 0.77 and 0.79 respectively. Whereas variance of C4se (the lowest mean) was the second largest at 1.02. C2e had the least variance (0.44) and C3e had the most (1.59).

Table 5.1.1. Maths classroom variables at time 2 for school 1 (Russian sample): Means, standard deviation (SD), and N with ANOVA results by classroom without controlling for prior achievement

Construct	C1e	C2e	C3e	C4se	C5se	C6se	C7se	C8se	*p*	η_p^2
Maths Year 5 school achievement	0.03 (1.13) n=20	-0.03 (1.05) n=20	-0.28 (0.77) n=14	-0.20 (0.97) n=23	-0.14 (1.06) n=22	0.59 (0.87) n=28	-0.03 (0.94) n=20	-0.09 (0.91) n=26	.071	.075
Maths performance	-0.14 (1.05) n=20	-0.23 (1.25) n=20	-0.09 (0.99) n=14	0.12 (0.87) n=23	-0.61 (0.98) n=27	0.68 (0.69) n=29	0.02 (0.99) n=23	-0.04 (0.70) n=29	**.000**	**.141**
Number line	-0.03 (0.86) n=20	0.34 (0.89) n=20	-0.14 (0.86) n=14	-0.52 (0.94) n=23	0.32 (0.68) n=27	-0.23 (0.93) n=28	-0.21 (0.71) n=23	0.60 (0.62) n=28	**.000**	**.173**
Maths self-perceived ability	0.24 (1.06) n=19	-0.05 (0.79) n=20	-0.19 (1.05) n=13	0.01 (0.90) n=22	-0.30 (0.77) n=27	0.26 (1.08) n=28	0.37 (0.84) n=23	-0.12 (1.20) n=27	.199	.055
Maths enjoyment	0.28 (1.18) n=19	0.03 (0.57) n=20	-0.29 (1.13) n=13	0.02 (0.84) n=23	-0.17 (0.87) n=27	0.03 (1.10) n=29	0.05 (1.10) n=23	0.12 (1.03) n=29	.792	.022
Maths classroom environment	-0.14 (0.86) n=21	0.46 (0.67) n=20	0.43 (0.99) n=14	0.28 (0.85) n=22	-0.16 (1.13) n=27	0.32 (0.80) n=29	0.23 (0.81) n=23	-0.01 (0.82) n=29	.089	.067
Maths classroom Student-teacher relations	-0.04 (0.84) n=21	0.48 (0.69) n=20	0.38 (1.05) n=14	0.20 (0.86) n=22	-0.26 (1.10) n=27	0.30 (0.85) n=29	0.24 (0.82) n=23	-0.03 (0.82) n=29	.083	.068
Maths classroom Peer competition	-0.30 (0.96) n=20	0.07 (0.70) n=20	0.34 (0.65) n=14	0.42 (0.65) n=22	0.33 (0.87) n=27	0.19 (0.69) n=29	0.04 (0.89) n=23	0.20 (0.86) n=28	.119	.063

Significant results in bold following a Bonferroni multiple testing correction of $p \leq .000$ ($p = .05$ divided by number of measures (k=114) across maths and geography, time 1, 2 & 3).

Table 5.1.1. Continued. Maths classroom variables at time 2 for school 1 (Russian sample): Means, standard deviation (SD), and N with ANOVA results by classroom without controlling for prior achievement

Construct	C1e	C2e	C3e	C4se	C5se	C6se	C7se	C8se	*p*	η_p^2
Maths classroom chaos	-0.28 (0.82) n=21	0.35 (0.66) n=20	0.51 (1.26) n=14	-0.56 (1.01) n=23	-0.55 (0.89) n=27	0.22 (0.80) n=29	0.64 (0.88) n=23	-0.25 (0.97) n=29	**.000**	**.197**
Maths homework behaviour	-0.02 (1.09) n=21	0.24 (0.98) n=20	0.26 (0.98) n=13	-0.27 (0.87) n=23	0.14 (1.08) n=27	-0.33 (1.08) n=29	-0.40 (1.01) n=23	-0.22 (1.17) n=29	.224	.051
Maths homework feedback	-0.66 (0.94) n=21	0.46 (0.57) n=19	-0.28 (1.02) n=13	-0.28 (0.73) n=23	0.24 (1.23) n=26	-0.26 (0.97) n=29	0.55 (0.86) n=23	0.15 (1.04) n=29	**.000**	**.146**
Maths homework total scale	-0.51 (0.89) n=21	0.23 (0.67) n=20	-0.32 (0.99) n=13	-0.07 (0.63) n=23	0.09 (1.18) n=27	-0.10 (1.13) n=29	0.53 (0.85) n=23	0.26 (0.94) n=28	.012	.096
Maths environment	-0.18 (0.96) n=20	0.04 (0.78) n=20	0.27 (0.97) n=13	-0.02 (0.98) n=23	-0.31 (1.14) n=27	0.56 (0.67) n=28	0.04 (1.05) n=23	-0.58 (0.82) n=27	.001	.127
Maths usefulness	-0.01 (0.80) n=19	-0.02 (0.71) n=20	-0.34 (0.62) n=13	-0.20 (0.69) n=23	0.22 (1.07) n=26	-0.10 (0.92) n=28	-0.60 (0.72) n=22	-0.10 (0.95) n=25	.077	.072
Maths anxiety	-0.15 (0.88) n=20	0.00 (0.80) n=19	-0.10 (1.10) n=14	0.20 (0.95) n=23	0.08 (1.14) n=24	-0.04 (1.02) n=27	0.06 (0.93) n=23	0.13 (1.03) n=23	.953	.013

Significant results in bold following a Bonferroni multiple testing correction of $p \leq .000$ (p = .05 divided by number of measures (k=114) across maths and geography, time 1, 2 & 3)

Maths homework feedback time 2. Modest significant differences between classrooms were found for homework feedback, $F(7,175) = 4.280$, $p < .001$, $\eta_p^2 = .146$, with the highest mean score revealed for C7se and the lowest for C1e. No significant pairwise comparisons were observed following multiple testing correction of $p \leq .000$ ($p = .05/114$). The difference between the highest and lowest mean fell just below significance ($p = .002$). However, Levene's test showed unequal variances for these analyses ($p = .027$), with the least variance for C2e (0.32) and the most for C5se (1.51).

School 2. ANOVA results by maths classroom at time 2,without controlling for prior achievement, are presented in Table 5.1.2. The results show for the majority of measures, no significant differences between maths classrooms following multiple testing correction of $p \leq .000$ ($p = .05/114$). Student-teacher relations and classroom chaos fell just below significance ($p = .001$ and $p = .009$ respectively). Levene's tests revealed that equal variances were assumed for most measures except classroom environment, homework feedback and maths anxiety (see Appendix 5, Table 5.1.2). For the following measure, significant average and variance differences between the classes were observed:

Maths classroom environment time 2. Moderate significant differences between classrooms were found for classroom environment , $F(2,33) = 13.456$, $p < .001$, $\eta_p^2 = .449$, with the highest mean score revealed for C10ce and the lowest for C11ce. Pairwise comparisons revealed that students in C10ce rated their classroom environment, on average higher than in C9ce and C11ce ($p < .001$), following multiple testing correction of $p \leq .000$ ($p = .05/114$). However, Levene's test showed unequal variances for these analyses ($p = .050$) with the most variance for C9ce (0.96) and the least variance for C10ce (0.20).

Table 5.1.2. Maths classroom variables at time 2 for school 2 (Russian sample): Means, standard deviation (SD), and N with ANOVA results by classroom without controlling for prior achievement

Construct	C9ce	C10ce	C11ce	*p*	η_p^2	Construct	C9ce	C10ce	C11ce	*p*	η_p^2
Maths Year 5 school achievement	0.28 (0.97) n=14	0.20 (1.12) n=9	-0.94 (0.69) n=7	.027	.235						
Maths performance	0.45 (0.98) n=15	0.18 (0.70) n=10	0.02 (0.93) n=11	.460	.046	Maths classroom chaos	0.00 (1.04) n=15	0.97 (0.66) n=10	-0.32 (1.02) n=12	.009	.242
Number line	-1.07 (0.93) n=14	-0.34 (1.21) n=10	0.29 (0.62) n=12	.003	.296	Maths homework behaviour	-0.16 (0.93) n=15	0.05 (0.77) n=10	0.35 (0.99) n=12	.360	.058
Maths self-perceived ability	-0.01 (0.86) n=15	-0.13 (0.85) n=10	0.05 (1.27) n=11	.925	.005	Maths homework feedback	-0.01 (1.02) n=14	0.34 (0.53) n=10	-0.18 (1.18) n=12	.462	.046
Maths enjoyment	0.19 (1.12) n=15	-0.08 (0.55) n=9	-0.45 (1.38) n=11	.358	.062	Maths homework total scale	0.13 (0.98) n=14	0.27 (0.53) n=10	-0.32 (1.09) n=12	.291	.072
Maths classroom environment	-0.88 (0.98) n=15	0.48 (0.45) n=10	-1.40 (0.94) n=11	**.000**	**.449**	Maths environment	0.12 (1.08) n=15	0.81 (0.96) n=9	-0.17 (1.16) n=12	.127	.118
Maths classroom student-teacher relations	-0.75 (1.03) n=15	0.49 (0.53) n=10	-1.12 (1.03) n=10	.001	.351	Maths usefulness	0.32 (1.08) n=15	-0.37 (0.45) n=10	0.62 (1.08) n=12	.063	.150
Maths classroom Peer competition	-0.79 (1.02) n=15	0.11 (0.61) n=10	-0.25 (1.08) n=9	.071	.157	Maths anxiety	-0.20 (0.76) n=15	-0.45 (0.70) n=10	-0.33 (1.26) n=12	.799	.013

Significant results in bold following a Bonferroni multiple testing correction of $p \leq .000$ ($p = .05$ divided by number of measures (k=114) across maths and geography, time 1, 2 & 3).

Differences between geography classrooms at time 2

School 1. ANOVA results by geography classroom at time 2,without controlling for prior achievement, are presented in Table 5.1.3. The results show for the majority of measures, no significant differences between geography classrooms following multiple testing correction of $p \leq .000$ ($p = .05/114$). Three measures that fell just below significance were Year 5 school achievement ($p = .009$), classroom environment ($p = .002$), and homework feedback ($p = .002$). Levene's tests revealed that equal variances were assumed for most measures except student-teacher relations and homework feedback (see Appendix 5, Table 5.1.3). For the following measures, significant average and variance differences between the classes were observed:

Geography performance time 2. Moderate significant differences between classrooms were found for geography performance, $F(7,173) = 6.227$, $p < .001$, $\eta_p^2 = .201$, with the highest mean score shown for C7se and the lowest for C5se. Pairwise comparisons revealed that students in C7se performed on average significantly higher than students in C5se and C1e ($p < .001$), following multiple testing correction of $p \leq .000$ ($p = .05/114$). Levene's test showed equal variances were assumed for these analyses ($p = .637$).

***Geography student-teacher relations time 2*.** A modest effect of classroom was found, $F(7,173) = 4.287$, $p < .001$, $\eta_p^2 = .148$, with the highest mean score shown for C3e and the lowest for C5se. No significant pairwise comparisons were observed following multiple testing correction of $p \leq .000$ ($p = .05/114$). However, Levene's test showed unequal variances for these analyses ($p = .016$) with the least variance for C2e (0.40) and the most for C4se (1.35).

Table 5.1.3. Geography classroom variables at time 2 for school 1 (Russian sample): Means, standard deviation (SD), and N with ANOVA results by classroom without controlling for prior achievement

Construct	C1e	C2e	C3e	C4se	C5se	C6se	C7se	C8se	*p*	η_p^2
Geography Year 5 school achievement	0.26 (0.67) n=20	-0.02 (1.02) n=20	-0.17 (0.80) n=14	-0.10 (1.18) n=23	-0.55 (0.90) n=22	0.55 (0.83) n=28	0.12 (1.00) n=20	0.05 (0.96) n=26	.009	.106
Geography performance	-0.32 (0.76) n=21	-0.26 (0.86) n=20	-0.10 (0.84) n=14	-0.04 (1.10) n=22	-0.43 (0.83) n=26	0.59 (0.88) n=28	0.81 (0.63) n=22	0.08 (0.91) n=28	**.000**	**.201**
Geography self-perceived ability	0.03 (0.89) n=20	-0.22 (0.81) n=20	0.52 (0.90) n=13	0.07 (1.04) n=22	-0.20 (1.14) n=25	-0.13 (1.08) n=28	0.43 (0.79) n=23	-0.51 (1.28) n=26	.029	.087
Geography enjoyment	0.27 (0.97) n=20	-0.13 (0.79) n=19	0.27 (0.93) n=14	0.14 (1.11) n=23	-0.22 (1.13) n=26	-0.18 (1.07) n=29	0.25 (0.88) n=23	-0.35 (1.16) n=26	.240	.051
Geography classroom environment	-0.16 (1.06) n=20	0.46 (0.67) n=20	0.48 (0.94) n=14	-0.25 (1.16) n=22	-0.39 (1.13) n=27	0.51 (0.77) n=29	0.04 (0.89) n=22	-0.13 (0.82) n=27	.002	.119
Geography classroom Student-teacher	-0.15 (1.02) n=20	0.46 (0.63) n=20	0.51 (0.87) n=14	-0.20 (1.16) n=22	-0.59 (1.13) n=27	0.49 (0.82) n=29	0.10 (0.86) n=22	-0.13 (0.77) n=27	**.000**	**.148**
Geography classroom peer competition	-0.16 (1.06) n=20	0.26 (0.81) n=20	0.19 (0.84) n=14	-0.28 (1.24) n=22	0.32 (1.01) n=26	0.36 (0.88) n=29	-0.12 (0.89) n=22	-0.16 (1.03) n=28	.134	.061

Significant results in bold following a Bonferroni multiple testing correction of $p \leq .000$ ($p = .05$ divided by number of measures (k=114) across maths and geography, time 1, 2 & 3).

Table 5.1.3. Continued. Geography classroom variables at time 2 for school 1 (Russian sample): Means, standard deviation (SD), and N with ANOVA results by classroom without controlling for prior achievement

Construct	C1e	C2e	C3e	C4se	C5se	C6se	C7se	C8se	*p*	η_p^2
Geography classroom chaos	-0.38 (0.96) n=20	0.16 (0.78) n=19	0.25 (1.16) n=14	-0.08 (1.09) n=22	-0.59 (0.95) n=27	0.06 (0.84) n=29	0.35 (0.76) n=21	-0.18 (0.94) n=28	.012	.097
Geography homework behaviour	-0.10 (1.08) n=20	0.30 (0.93) n=20	0.09 (1.01) n=14	-0.27 (0.87) n=22	0.23 (1.10) n=26	-0.18 (1.03) n=29	-0.45 (1.00) n=23	0.27 (1.06) n=27	.100	.066
Geography homework feedback	-0.77 (0.60) n=20	0.24 (0.71) n=20	0.19 (1.07) n=14	-0.41 (1.11) n=22	-0.32 (1.16) n=26	0.07 (0.97) n=29	0.34 (1.03) n=23	0.17 (0.93) n=27	.002	.123
Geography homework total scale	-0.53 (0.66) n=20	0.05 (0.82) n=20	0.12 (1.15) n=14	-0.16 (0.93) n=22	-0.38 (1.23) n=26	0.15 (1.00) n=29	0.43 (0.92) n=23	-0.02 (0.91) n=27	.032	.084
Geography environment	-0.27 (0.93) n=20	-0.25 (0.88) n=18	-0.03 (1.00) n=13	-0.30 (0.69) n=21	-0.38 (0.91) n=25	0.98 (0.89) n=29	-0.08 (0.99) n=22	-0.27 (1.03) n=25	**.000**	**.213**
Geography usefulness	0.14 (0.83) n=20	-0.07 (0.60) n=20	0.17 (0.81) n=14	-0.40 (1.28) n=23	-0.26 (1.11) n=27	0.13 (0.92) n=29	0.13 (1.12) n=23	0.10 (0.88) n=27	.351	.043
Geography anxiety	-0.29 (0.96) n=20	-0.11 (0.90) n=20	-0.07 (1.17) n=13	0.34 (1.02) n=23	-0.07 (1.04) n=27	0.00 (0.88) n=28	-0.20 (0.75) n=23	0.15 (1.06) n=27	.457	.038

Significant results in bold following a Bonferroni multiple testing correction of $p \leq .000$ ($p = .05$ divided by number of measures (k=114) across maths and geography, time 1, 2 & 3).

Table 5.1.4. Geography classroom variables at time 2 for school 2 (Russian sample): Means, standard deviation (SD), and N with ANOVA results by classroom without controlling for prior achievement

Construct	C9ce	C10ce	C11ce	*p*	η_p^2	Construct	C9ce	C10ce	C11ce	*p*	η_p^2
Geography Year 5 school achievement	0.23 (0.81) n=14	0.41 (0.86) n=9	-0.45 (1.31) n=7	.189	.116	Geography classroom chaos	0.23 (1.03) n=15	1.21 (0.64) n=10	-0.19 (1.08) n=12	.006	.264
Geography performance	0.00 (0.72) n=15	0.11 (0.67) n=10	-0.15 (0.81) n=11	.721	.020	Geography homework behaviour	-0.45 (0.80) n=15	0.00 (0.80) n=9	0.38 (1.27) n=12	.104	.128
Geography self-perceived ability	0.28 (0.80) n=15	-0.02 (0.60) n=10	0.32 (0.69) n=11	.495	.042	Geography homework feedback	0.45 (0.93) n=15	0.13 (0.67) n=9	0.22 (0.99) n=12	.664	.025
Geography enjoyment	0.38 (0.89) n=15	-0.04 (0.77) n=10	-0.06 (0.76) n=12	.303	.068	Geography homework total scale	0.60 (0.91) n=15	0.18 (0.68) n=9	-0.07 (1.24) n=12	.212	.090
Geography classroom environment	-0.33 (0.91) n=15	0.13 (0.85) n=10	-0.09 (1.08) n=12	.490	.041	Geography environment	0.02 (0.94) n=15	0.45 (0.62) n=9	0.25 (1.14) n=12	.546	.036
Geography classroom Student-teacher	-0.20 (0.94) n=15	0.21 (0.84) n=10	-0.12 (1.18) n=12	.593	.030	Geography usefulness	0.60 (0.92) n=15	-0.12 (0.91) n=10	-0.05 (1.05) n=11	.123	.119
Geography classroom peer competition	-0.52 (1.04) n=15	-0.12 (0.94) n=10	0.00 (1.02) n=12	.383	.055	Geography anxiety	-0.22 (0.78) n=14	-0.16 (0.71) n=10	-0.07 (0.90) n=11	.907	.006

Significant results in bold following a Bonferroni multiple testing correction of $p \leq .000$ (p = .05 divided by number of measures (k=114) across maths and geography, time 1, 2 & 3).

***Geography environment time 2*.** A moderate effect of classroom was found, $F(7,165) = 6.372$, $p < .001$, $\eta_p^2 = .213$, with the highest mean score shown for C6se and the lowest for C5se. Pairwise comparisons revealed that students in C6se rated their geography environment in the use of equipment such as compasses etc., more highly than students in C5se and C4se ($p < .001$), following multiple testing correction of $p \leq .000$ ($p = .05/114$). Levene's test showed equal variances were assumed for these analyses ($p = .657$).

School 2. ANOVA results by geography classroom at time 2,without controlling for prior achievement, are presented in Table 5.1.4. The results show no significant differences between geography classrooms following multiple testing correction of $p \leq .000$ ($p = .05/114$). Geography classroom chaos fell just below significance ($p = .006$). Levene's tests revealed that equal variances were assumed for most measures except homework behaviour (see Appendix 5, Table 5.1.4).

Maths and geography teacher group differences at time 2

Further analyses were conducted to establish whether patterns of significant differences across maths and geography teacher groups found at time 1 persisted at time 2,without controlling for primary school achievement. The eleven classes across the two schools were grouped by maths teacher (six teachers across eleven classes), and by geography teacher (five teachers across eleven classes). Appendix 5, Table 5.1.7 presents each class and their corresponding teachers.

ANOVA results at time 2, without controlling for prior achievement, can be found for maths teachers in Table 5.1.5, and for geography teachers in Table 5.1.6. The results show for the majority of the measures, no significant differences between maths or geography teacher groupings following multiple testing correction of $p \leq .001$ ($p = .05/57$). Measures just below the significance threshold were: maths student-teacher relations ($p = .002$), and maths environment ($p = .004$). Levene's tests revealed that equal variances were assumed for most measures, except maths classroom chaos and geography self-perceived ability (see Tables 5.1.5 and 5.1.6). For the following measures, significant average and variance differences between the classes were observed:

***Maths performance time 2*.** Modest significant differences were found between teacher groupings for maths performance, $F(5,215) = 4.571$, $p = .001$, $\eta_p^2 = .096$, with students studying with teacher TM5 having the highest score for maths performance and students studying with teacher TM3 having the lowest. Following multiple testing correction of $p \leq .001$ ($p = .05/57$), pairwise comparisons revealed that students studying with teacher TM5 had on average, significantly higher maths performance than students studying with teacher TM6, but not teacher TM3 (lowest mean score). As Levene's test showed equal variances were assumed for these analyses ($p = .144$), it appears unusual that the highest (TM5) and lowest (TM3) means did not reach significance despite having the largest mean difference of 0.86. However, this pairwise comparison also had the largest standard error of 0.22 compared to the mean difference and standard error between the significant pair TM5 and TM6 of -0.81 (SE = 0.15).

Table 5.1.5. Maths teacher groups time 2 (Russian sample): Means, standard deviation (SD) and N for maths classroom variables with ANOVA results by teacher group, without controlling for prior achievement

Construct	TM1	TM2	TM3	TM4	TM5	TM6	*p*	η_p^2
Maths Year 5 school achievement	0.28 (0.97) n=14	-0.30 (1.09) n=16	0.00 (1.08) n=40	-0.28 (0.77) n=14	0.59 (0.87) n=28	-0.12 (0.96) n=91	.010	.074
Maths performance	0.45 (0.98) n=15	0.10 (0.81) n=21	-0.19 (1.14) n=40	-0.09 (0.99) n=14	0.68 (0.69) n=29	-0.14 (0.92) n=102	**.001**	**.096**
Number line	-1.07 (0.93) n=14	0.01 (0.96) n=22	0.16 (0.89) n=40	-0.14 (0.86) n=14	-0.23 (0.93) n=28	0.09 (0.85) n=101	**.000**	**.103**
Maths self-perceived ability	-0.01 (0.86) n=15	-0.04 (1.07) n=21	0.09 (0.93) n=39	-0.19 (1.05) n=13	0.26 (1.08) n=28	-0.03 (0.97) n=99	.726	.013
Maths enjoyment	0.19 (1.12) n=15	-0.28 (1.08) n=20	0.15 (0.92) n=39	-0.29 (1.13) n=13	0.03 (1.10) n=29	0.01 (0.96) n=102	.529	.019
Maths classroom environment	-0.88 (0.98) n=15	-0.51 (1.20) n=21	0.15 (0.82) n=41	0.43 (0.99) n=14	0.32 (0.80) n=29	0.07 (0.92) n=101	**.000**	**.114**
Maths classroom student-teacher relations	-0.75 (1.03) n=15	-0.32 (1.15) n=20	0.21 (0.80) n=41	0.38 (1.05) n=14	0.30 (0.85) n=29	0.02 (0.92) n=101	.002	.083
Maths classroom peer competition	-0.79 (1.02) n=15	-0.06 (0.86) n=19	-0.11 (0.85) n=40	0.34 (0.65) n=14	0.19 (0.69) n=29	0.25 (0.83) n=100	**.000**	**.107**

Significant results in bold following a Bonferroni multiple testing correction of $p \leq .001$ ($p = .05$ divided by number of measures (k=57) across maths and geography, time 1, 2 & 3).

Table 5.1.5. Continued. Maths teacher groups time 2 (Russian sample): Means, standard deviation (SD) and N for maths classroom variables with ANOVA results by teacher group, without controlling for prior achievement

Construct	TM1	TM2	TM3	TM4	TM5	TM6	*p*	η_p^2
Maths classroom chaos	0.00 (1.04) n=15	0.26 (1.08) n=22	0.03 (0.80) n=41	0.51 (1.26) n=14	0.22 (0.80) n=29	-0.20 (1.04) n=102	.058	.048
Maths homework behaviour	-0.16 (0.93) n=15	0.21 (0.89) n=22	0.11 (1.03) n=41	0.26 (0.98) n=13	-0.33 (1.08) n=29	-0.18 (1.05) n=102	.206	.033
Maths homework feedback	-0.01 (1.02) n=14	0.05 (0.96) n=22	-0.13 (0.96) n=40	-0.28 (1.02) n=13	-0.26 (0.97) n=29	0.17 (1.02) n=101	.263	.030
Maths homework total scale	0.13 (0.98) n=14	-0.05 (0.91) n=22	-0.15 (0.86) n=41	-0.32 (0.99) n=13	-0.10 (1.13) n=29	0.20 (0.94) n=101	.215	.032
Maths environment	0.12 (1.08) n=15	0.25 (1.16) n=21	-0.07 (0.87) n=40	0.27 (0.97) n=13	0.56 (0.67) n=28	-0.24 (1.02) n=100	.004	.078
Maths usefulness	0.32 (1.08) n=15	0.17 (0.98) n=22	-0.02 (0.75) n=39	-0.34 (0.62) n=13	-0.10 (0.92) n=28	-0.15 (0.91) n=96	.255	.031
Maths anxiety	-0.20 (0.76) n=15	-0.39 (1.02) n=22	-0.08 (0.83) n=39	-0.10 (1.10) n=14	-0.04 (1.02) n=27	0.12 (1.00) n=93	.337	.027

Significant results in bold following a Bonferroni multiple testing correction of $p \leq .001$ ($p = .05$ divided by number of measures (k=57) across maths and geography, time 1, 2 & 3).

Table 5.1.6. Geography teacher groups time 2 (Russian sample): Means, standard deviation (SD) and N for geography classroom variables with ANOVA results by teacher group, without controlling for prior achievement

Construct	TG1	TG2	TG3	TG4	TG5	*p*	η_p^2
Geography Year 5 school achievement	0.13 (0.98) n=30	-0.12 (0.99) n=68	0.55 (0.83) n=28	0.12 (0.87) n=40	-0.13 (1.04) n=37	.023	.055
Geography performance	-0.01 (0.72) n=36	0.12 (0.94) n=76	0.59 (0.88) n=28	-0.29 (0.80) n=41	-0.06 (1.00) n=36	.002	.078
Geography self-perceived ability	0.21 (0.71) n=36	-0.11 (1.15) n=74	-0.13 (1.08) n=28	-0.09 (0.85) n=40	0.23 (1.00) n=35	.260	.025
Geography enjoyment	0.13 (0.82) n=37	-0.12 (1.09) n=75	-0.18 (1.07) n=29	0.07 (0.90) n=39	0.19 (1.04) n=37	.399	.019
Geography classroom environment	-0.13 (0.95) n=37	-0.17 (0.97) n=76	0.51 (0.77) n=29	0.15 (0.93) n=40	0.03 (1.13) n=36	.018	.054
Geography classroom Student-teacher relations	-0.06 (0.99) n=37	-0.23 (0.97) n=76	0.49 (0.82) n=29	0.15 (0.89) n=40	0.08 (1.10) n=36	.013	.057
Geography classroom peer competition	-0.24 (1.01) n=37	0.02 (1.00) n=76	0.36 (0.88) n=29	0.05 (0.95) n=40	-0.10 (1.11) n=36	.175	.029

Significant results in bold following a Bonferroni multiple testing correction of $p \leq .001$ ($p = .05$ divided by number of measures (k=57) across maths and geography, time 1, 2 & 3).

Table 5.1.6. Continued. Geography teacher groups time 2 (Russian sample): Means, standard deviation (SD) and N for geography classroom variables with ANOVA results by teacher group, without controlling for prior achievement

Construct	TG1	TG2	TG3	TG4	TG5	*p*	η_p^2
Geography classroom chaos	0.36 (1.09) n=37	-0.18 (0.96) n=76	0.06 (0.84) n=29	-0.12 (0.91) n=39	0.05 (1.12) n=36	.093	.037
Geography homework behaviour	-0.06 (1.02) n=36	0.04 (1.09) n=76	-0.18 (1.03) n=29	0.10 (1.02) n=40	-0.13 (0.93) n=36	.734	.009
Geography homework feedback	0.29 (0.88) n=36	0.05 (1.06) n=76	0.07 (0.97) n=29	-0.26 (0.83) n=40	-0.18 (1.12) n=36	.118	.034
Geography homework total scale	0.27 (1.01) n=36	-0.01 (1.07) n=76	0.15 (1.00) n=29	-0.24 (0.79) n=40	-0.05 (1.02) n=36	.214	.027
Geography environment	0.20 (0.94) n=36	-0.25 (0.97) n=72	0.98 (0.89) n=29	-0.26 (0.89) n=38	-0.20 (0.82) n=34	**.000**	**.180**
Geography usefulness	0.20 (0.99) n=36	-0.02 (1.04) n=77	0.13 (0.92) n=29	0.03 (0.72) n=40	-0.18 (1.14) n=37	.521	.015
Geography anxiety	-0.16 (0.78) n=35	-0.03 (0.97) n=77	0.00 (0.88) n=28	-0.20 (0.92) n=40	0.19 (1.08) n=36	.412	.018

Significant results in bold following a Bonferroni multiple testing correction of $p \leq .001$ ($p = .05$ divided by number of measures (k=57) across maths and geography, time 1, 2 & 3).

***Number line time 2*.** Small significant differences between teacher groupings were found for the number line task, $F(5,213) = 4.884$, $p < .001$, $\eta_p^2 = .103$, with the lowest (optimum) mean score shown for TM1 and the highest for TM3. However, following multiple testing correction of $p \leq .001$ ($p = .05/57$), pairwise comparisons revealed that students did not on average, perform number estimation significantly better or worse when being taught by a specific teacher. Levene's test showed equal variances were assumed for these analyses ($p = .948$).

***Maths classroom environment time 2*.** Modest significant differences between teacher groupings were found for classroom environment, $F(5,215) = 5.537$, $p = .001$, $\eta_p^2 = .114$, with the highest mean score shown for TM4 and the lowest for TM1. However, pairwise comparisons, following multiple testing correction of $p \leq .001$ ($p = .05/57$) showed that students did not on average, rate their classroom environments differently when taught by different teachers. Levene's test showed equal variances were assumed for these analyses ($p = .058$).

***Maths classroom peer competition time 2*.** Small but significant differences between teacher groupings were found for peer competition, $F(5,211) = 5.063$, $p = .001$, $\eta_p^2 = .107$, with the highest mean score shown for TM4 and the lowest for TM1. However, pairwise comparisons following multiple testing correction of $p \leq .001$ ($p = .05/57$) revealed that students on average, did not evaluate peer competition differently when taught by different teachers. Levene's test showed equal variances were assumed for these analyses ($p = .521$).

***Geography environment time 2*.** Modest significant differences between teacher groupings were found for geography environment, $F(5,204) =$

11.219, p = .001, η_p^2= .180, with the highest mean score shown for TG3 and the lowest for TG4. Pairwise comparisons revealed that students studying with teacher TG3 rated their geography environment in the use of equipment such as compasses etc., on average, more highly than students studying with teachers TG2, TG4 and TG5 ($p < .001$), following multiple testing correction of $p \leq .001$ (p = .05/57). Levene's test showed equal variances were assumed for these analyses (p = .850).

Class and teacher group ranking by mean score at time 2

Classes and teacher groups were ranked by their mean scores (highest to lowest) across measures that demonstrated a significant effect of class or teacher group. The expectation being that more consistency of ranking position across measures would indicate a stronger influence of class or teacher group. If the level of consistency was higher than consistency found at time 1 (Chapter 4, pp. 139, 140, 141, 154 and 155) this might indicate a stronger influence of current subject teacher as opposed to primary school.

Maths classroom. The results for school 1,without controlling for prior achievement, in Table 5.1.7 show very little consistency of rank across the measures for most classes. For some classes their ranks sit predominantly towards the lower ranks, for example, C5se, C8se and C1e. While others, for example C6se and C7se, sit towards the higher ranks. However, there is still some variation even for these classes, suggesting a weaker effect of classroom. Less correspondence is shown between time 1 and 2 rankings, especially between primary and year 5 school achievement. This may suggest a weakening of primary school influences for some classes, for example, classes C1e and C2e are ranking in first and second places, respectively for year 5 achievement, whereas they ranked seventh and fourth respectively for primary school achievement.

For school 2, in Table 5.1.8, the results show much more correspondence of rank across the measures, however, there are fewer classes to vary. Class C11ce is consistently in third place for all measures and C10ce appears most frequently in first place.

Geography classroom. The results for school 1, without controlling for prior achievement, in Table 5.1.9 show slightly more correspondence of rank across the measures with less variation for individual classrooms. For example, C5se ranks in eighth place for five out of six measures and C4se ranks seventh for four our six measures. Similarly to maths classroom, the same classes sit towards the upper and lower ranks, for example, C6se ranks first and second place for five out of six measures. The higher level of consistency suggests a slightly stronger effect of geography classroom across the measures. Similarly to maths classroom, there is less correspondence with time 1 (Chapter 4, p. 141) especially between primary and year 5 school achievement for the majority of classes. Class C6se, however remains in first place for both primary and year 5 achievement.

The results for school 2,without controlling for prior achievement, in Table 5.1.10 show complete consistency across the measures with C10ce in first, C9ce in second and C11ce in third place. However, the small number of classes may account for this to some extent.

Table 5.1.7. Maths classroom variables at time 2 for school 1 (Russian sample): Classrooms ranked by means (highest = 1 to lowest = 8) for measures demonstrating a significant effect of maths classroom, without controlling for prior achievement

Construct	C1e	C2e	C3e	C4se	C5se	C6se	C7se	C8se	*p*	η_p^2
Maths Year 5 school achievement**	2nd	3rd	8th	7th	6th	1st	4th	5th	.071	.075
Maths performance	6th	7th	5th	2nd	8th	1st	3rd	4th	**.000**	**.141**
Number line	5th	7th	4th	1st	6th	2nd	3rd	8th	**.000**	**.173**
Maths classroom chaos	6th	3rd	2nd	8th	7th	4th	1st	5th	**.000**	**.197**
Maths homework feedback	8th	2nd	7th	6th	3rd	5th	1st	4th	**.000**	**.146**
Maths Environment*	6th	4th	2nd	5th	7th	1st	3rd	8th	.001	.127

Significant results in bold following a Bonferroni multiple testing correction of $p \leq .000$ ($p = .05$ divided by number of measures (k=114) across maths and geography, time 1, 2 & 3) . *Just below significance and **not significant but ranked for comparison.

Table 5.1.8. Maths classroom variables at time 2 for school 2 (Russian sample): Classrooms ranked by means (highest = 1 to lowest = 3) for measures demonstrating a significant effect of maths classroom, without controlling for prior achievement

Construct	C9ce	C10ce	C11ce	*p*	η_p^2
Maths Year 5 school achievement**	**1st**	**2nd**	**3rd**	.027	.235
Number line*	**1st**	**2nd**	**3rd**	.003	.296
Maths classroom environment	**2nd**	**1st**	**3rd**	**.000**	**.449**
Maths classroom student-teacher relations*	**2nd**	**1st**	**3rd**	.001	.351
Maths classroom chaos*	**2nd**	**1st**	**3rd**	.009	.242

Significant results in bold following a Bonferroni multiple testing correction of $p \leq .000$ (p = .05 divided by number of measures (k=114) across maths and geography, time 1, 2 & 3) . *Just below significance and **not significant but ranked for comparison

Table 5.1.9. Geography classroom variables at time 2 for school 1 (Russian sample): Classrooms ranked by means (highest = 1 to lowest = 8) for measures demonstrating a significant effect of geography classroom, without controlling for prior achievement

Construct	C1e	C2e	C3e	C4se	C5se	C6se	C7se	C8se	*p*	η_p^2
Geography Year 5 school achievement*	2nd	5th	7th	6th	8th	1st	3rd	4th	.009	.106
Geography performance	7th	6th	5th	4th	8th	2nd	1st	3rd	**.000**	**.201**
Geography classroom environment*	6th	3rd	2nd	7th	8th	1st	4th	5th	.002	.119
Geography classroom Student-teacher	6th	3rd	1st	7th	8th	2nd	4th	5th	**.000**	**.148**
Geography homework Feedback*	8th	2nd	3rd	7th	6th	5th	1st	4th	.002	.123
Geography environment	6th	4th	2nd	7th	8th	1st	3rd	5th	**.000**	**.213**

Significant results in bold following a Bonferroni multiple testing correction of $p \leq .000$ ($p = .05$ divided by number of measures (k=114) across maths and geography, time 1, 2 & 3). *Just below significance but ranked for comparison.

Table 5.1.10. Geography classroom variables at time 2 for school 2 (Russian sample): Classrooms ranked by means (highest = 1 to lowest = 3) for measures demonstrating a significant effect of geography classroom, without controlling for prior achievement

Construct	C9ce	C10ce	C11ce	*p*	η_p^2
Geography Year 5 school achievement**	**2nd**	**1st**	**3rd**	.189	.116
Geography classroom Chaos*	**2nd**	**1st**	**3rd**	.006	.264

Significant results in bold following a Bonferroni multiple testing correction of $p \leq .000$ ($p = .05$ divided by number of measures (k=114) across maths and geography, time 1, 2 & 3). *Just below significance and **not significant but ranked for comparison.

Table 5.1.11. Maths Teacher groups at time 2 (Russian sample): ranked by means (highest = 1 to lowest = 6) for measures demonstrating a significant effect of maths teacher without controlling for prior achievement

Construct	TM1	TM2	TM3	TM4	TM5	TM6	*p*	η_p^2
Maths performance	2nd	3rd	6th	4th	1st	5th	**.001**	**.096**
Number line	1st	4th	6th	3rd	2nd	5th	**.000**	**.103**
Maths classroom environment	6th	5th	3rd	1st	2nd	4th	**.000**	**.114**
Maths classroom student-teacher relations*	6th	5th	3rd	1st	2nd	4th	.002	.083
Maths classroom peer competition	6th	4th	5th	1st	3rd	2nd	**.000**	**.107**
Maths environment*	4th	3rd	5th	2nd	1st	6th	.004	.078

Significant results in bold following a Bonferroni multiple testing correction of $p \leq .001$ ($p = .05$ divided by number of measures (k=57) across maths and geography, time 1, 2 & 3) . *Just below significance and but ranked for comparison

Table 5.1.12. Geography Teacher groups at time 2 (Russian sample): ranked by means (highest = 1 to lowest = 5) for measures demonstrating a significant effect of geography teacher without controlling for prior achievement

Construct	TG1	TG2	TG3	TG4	TG5	*p*	η_p^2
Geography performance*	**3rd**	**2nd**	**1st**	**5th**	**4th**	.002	.078
Geography environment	**2nd**	**4th**	**1st**	**5th**	**3rd**	**.000**	**.180**

Significant results in bold following aBonferroni multiple testing correction of $p \leq .001$($p = .05$ divided by number of measures (k=57) acrossmaths and geography, time 1, 2 & 3). *Just below significance and but ranked for comparison.

Maths Teacher. The results, without controlling for prior achievement, in Table 5.1.11 show some correspondence of rank across measures for some teacher groups. As with classrooms, some groups sit towards the upper ranks, for example, TM4 and TM5, while others, such as TM1 and TM6 sit towards the lower ranks. There is some variation but less than for classrooms. The rankings show some correspondence with time 1 (Chapter 4, p.154) especially for classroom environment and student-teacher relations. However, only two teacher groups, TM1 and TM5, are consistent for maths performance across time 1 and 2.

Geography Teacher. The results, without controlling for prior achievement, in Table 5.1.12 show ranking for just two measures due to very few measures reaching significance. Some consistency is shown with four out of five groups showing complete correspondence, or very close position of rank. There is also some consistency with time 1 (Chapter 4, p.155).

Differences between maths classes at time 3

School 1. ANOVA results by maths classroom at time 3 for school 1, without controlling for prior achievement, are presented in Table 5.1.13. The results show for the majority of measures, no significant differences between maths classrooms following multiple testing correction of $p \leq .000$ ($p = .05/114$). Maths performance ($p = .002$), and classroom chaos ($p =.001$) fell just below significance. Levene’s tests revealed that equal variances were assumed for most measures, except peer competition (see Appendix 5, Table 5.1.1). The following two measures were significant:

Table 5.1.13. Maths classroom variables at time 3 for school 1 (Russian sample): Means, standard deviation (SD), and N with ANOVA results by classroom without controlling for prior achievement

Construct	C1e	C2e	C3e	C4se	C5se	C6se	C7se	C8se	*p*	η_p^2
Maths performance	0.31 (0.55) n=21	0.07 (1.06) n=20	-0.22 (0.95) n=11	-0.12 (0.93) n=25	-0.65 (0.97) n=19	0.53 (0.86) n=29	-0.28 (1.00) n=22	-0.03 (1.07) n=26	.002	.127
Number line	0.23 (0.84) n=24	0.14 (0.90) n=22	-0.02 (0.92) n=11	-0.03 (1.16) n=24	0.54 (0.62) n=20	-0.16 (0.89) n=29	0.00 (0.93) n=21	0.29 (0.86) n=26	.211	.055
Maths self-perceived ability	0.16 (0.98) n=23	-0.30 (0.84) n=22	0.09 (0.90) n=11	-0.19 (0.92) n=25	0.05 (1.05) n=20	0.06 (1.06) n=28	0.25 (1.02) n=22	-0.04 (1.10) n=26	.653	.029
Maths enjoyment	0.06 (0.99) n=22	0.05 (0.79) n=19	0.33 (0.85) n=11	-0.12 (0.81) n=25	-0.28 (0.75) n=19	0.09 (1.24) n=27	0.37 (0.87) n=22	-0.22 (1.27) n=25	.373	.045
Maths classroom environment	0.10 (0.79) n=23	0.52 (0.61) n=22	0.66 (0.69) n=11	0.09 (0.64) n=25	-0.36 (0.80) n=19	0.47 (0.70) n=29	0.00 (0.61) n=22	0.05 (0.90) n=25	**.000**	**.142**
Maths classroom Student-teacher relations	0.05 (0.77) n=23	0.67 (0.60) n=22	0.68 (0.79) n=11	-0.06 (0.64) n=25	-0.44 (0.75) n=19	0.50 (0.81) n=29	0.01 (0.63) n=22	0.07 (0.86) n=25	**.000**	**.187**
Maths classroom Peer competition	0.02 (0.89) n=23	-0.10 (0.92) n=22	0.35 (0.58) n=11	0.49 (0.73) n=25	-0.17 (1.18) n=20	0.29 (0.74) n=29	-0.07 (0.94) n=22	0.15 (0.89) n=26	.141	.062

Significant results in bold following a Bonferroni multiple testing correction of $p \leq .000$ (p = .05 divided by number of measures (k=114) across maths and geography, time 1, 2 & 3).

Table 5.1.13. Continued. Maths classroom variables at time 3 for school 1 (Russian sample): Means, standard deviation (SD), and N with ANOVA results by classroom without controlling for prior achievement

Construct	C1e	C2e	C3e	C4se	C5se	C6se	C7se	C8se	*p*	η_p^2
Maths classroom chaos	-0.32 (1.08) n=24	0.65 (0.86) n=22	0.10 (1.32) n=11	-0.29 (0.87) n=25	-0.30 (0.95) n=20	0.24 (0.98) n=29	0.47 (0.94) n=22	-0.33 (0.90) n=27	.001	.134
Maths homework behaviour	-0.24 (0.89) n=23	0.39 (0.81) n=21	0.51 (0.51) n=11	0.36 (1.08) n=25	-0.14 (0.98) n=20	-0.35 (0.97) n=29	-0.31 (1.07) n=21	0.04 (1.16) n=27	.022	.091
Maths homework feedback	0.01 (0.92) n=23	0.50 (0.82) n=21	-0.30 (0.90) n=11	-0.38 (0.89) n=25	-0.56 (0.81) n=20	0.10 (1.18) n=29	0.18 (0.88) n=20	0.19 (1.18) n=27	.012	.100
Maths homework total scale	0.20 (0.66) n=23	0.24 (0.79) n=21	-0.35 (0.76) n=11	-0.34 (0.91) n=24	-0.28 (0.89) n=20	0.26 (1.15) n=29	0.30 (0.98) n=21	0.23 (1.01) n=26	.050	.080
Maths environment	0.10 (1.10) n=23	0.53 (0.61) n=22	0.05 (1.09) n=11	-0.06 (0.99) n=25	-0.10 (0.96) n=20	0.40 (0.91) n=28	-0.19 (0.94) n=22	-0.45 (0.91) n=24	.011	.102
Maths usefulness	-0.27 (0.80) n=22	-0.18 (0.70) n=22	-0.01 (0.94) n=9	-0.26 (0.86) n=24	0.24 (1.05) n=20	-0.17 (0.92) n=28	-0.60 (0.89) n=22	0.19 (1.06) n=27	.062	.077
Maths anxiety	-0.10 (0.75) n=22	0.33 (0.92) n=22	-0.30 (0.77) n=11	0.40 (0.99) n=25	-0.26 (1.00) n=20	-0.16 (0.98) n=27	0.02 (0.99) n=22	0.13 (1.08) n=26	.159	.060

Significant results in bold following a Bonferroni multiple testing correction of $p \leq .000$ ($p = .05$ divided by number of measures (k=114) across maths and geography, time 1, 2 & 3).

Maths classroom environment time 3. Modest significant differences between classrooms were found for classroom environment, $F(7,168) = 3.966$, $p < .001$, $\eta_p^2 = .142$, with the highest mean score revealed for C3e and the lowest for C5se. Following multiple testing correction of $p \leq .000$ ($p = .05/114$), pairwise comparisons revealed that students on average, did not evaluate their classroom environments differently across class groups. Levene's test showed equal variances were assumed for these analyses ($p = .205$).

***Maths student-teacher relations time 3*.** Modest significant differences between classrooms were found for student-teacher relations, $F(7,168) = 5.533$, $p < .001$, $\eta_p^2 = .187$, with the highest mean score revealed for C3e and the lowest for C5se. Following multiple testing correction of $p \leq .000$ ($p = .05/114$), pairwise comparisons revealed that on average, class C5se rated their student-teacher relations significantly lower than class C2e, but not C3e. As Levene's test showed equal variances were assumed for these analyses ($p = .232$), it appears unusual that the significant difference did not occur between the highest and lowest means. However, while the mean difference between C5se and C3e (-1.12) was close in size to that between C5se and C2e (1.11) the standard error of 0.29 was larger than that between C5se and C2e (0.21).

School 2. ANOVA results by maths classroom at time 3 for school 2, without controlling for prior achievement, are presented in Table 5.1.14. The results show for the majority of measures, no significant differences between maths classrooms following multiple testing correction of $p \leq .000$ ($p = .05/114$). Student-teacher relations, peer competition and classroom chaos ($p = .001$) fell just below significance. Levene's tests revealed that equal variances were assumed for most measures except classroom chaos, homework feedback, homework total scale and maths anxiety (see Appendix 5, Table 5.1.2). The following two measures were significant:

***Maths classroom environment time 3*.** Moderate significant differences between classrooms were found for classroom environment, $F(2,32) = 15.703$, $p < .001$, $\eta_p^2 = .495$, with the highest mean score revealed for C10ce and the lowest for C11ce. Following multiple testing correction of $p \leq .000$ ($p = .05/114$), pairwise comparisons showed that, on average, students in class C10ce rated their classroom environment significantly higher than students in class C11ce. Levene's test showed equal variances were assumed for these analyses ($p = .428$).

***Maths homework behaviour time 3*.** Moderate significant differences between classrooms were found, $F(2,35) = 11.437$, $p < .001$, $\eta_p^2 = .395$, with the highest mean score revealed for C9ce and the lowest for C11ce. Following multiple testing correction of $p \leq .000$ ($p = .05/114$), pairwise comparisons revealed that students in C9ce rated their homework behaviour significantly higher than students in class C11ce ($p < .000$). Levene's test showed equal variances were assumed for these analyses ($p = .282$).

Table 5.1.14. Maths classroom variables at time 3 for school 2 (Russian sample): Means, standard deviation (SD), and N with ANOVA results by classroom without controlling for prior achievement

Construct	C9ce	C10ce	C11ce	*p*	η_p^2	Construct	C9ce	C10ce	C11ce	*p*	η_p^2
Maths performance	0.38 (0.91) n=16	0.45 (0.72) n=11	-0.17 (1.30) n=11	.272	.072	Maths classroom chaos	0.02 (0.87) n=16	0.79 (0.44) n=11	-0.45 (0.66) n=11	.001	.329
Number line	-0.55 (1.14) n=15	-0.01 (0.79) n=11	0.81 (1.04) n=11	.008	.250	Maths homework behaviour	-0.10 (0.87) n=16	0.24 (0.63) n=11	1.26 (0.60) n=11	**.000**	**.395**
Maths self-perceived ability	-0.34 (0.86) n=16	0.31 (0.85) n=10	0.47 (1.28) n=9	.103	.132	Maths homework feedback	0.03 (0.61) n=16	0.23 (0.64) n=11	0.17 (1.53) n=11	.870	.008
Maths enjoyment	-0.08 (1.13) n=16	-0.13 (0.68) n=9	0.48 (1.07) n=9	.355	.065	Maths homework total scale	0.16 (0.69) n=16	0.18 (0.53) n=11	-0.61 (1.25) n=11	.051	.156
Maths classroom environment	-0.39 (0.93) n=16	0.56 (0.63) n=11	-1.55 (0.76) n=8	**.000**	**.495**	Maths environment	0.37 (0.96) n=15	0.27 (0.87) n=11	-0.56 (1.18) n=10	.070	.149
Maths classroom student-teacher relations	-0.42 (0.98) n=16	0.47 (0.64) n=11	-1.29 (0.92) n=8	.001	.373	Maths usefulness	0.05 (0.91) n=16	0.32 (0.69) n=11	0.92 (1.06) n=11	.059	.149
Maths classroom Peer competition	-0.25 (1.00) n=16	0.29 (0.61) n=11	-1.50 (1.310 n=10	.001	.339	Maths anxiety	-0.15 (0.95) n=15	-0.73 (0.59) n=11	-0.33 (1.07) n=10	.273	.076

Significant results in bold following a Bonferroni multiple testing correction of $p \leq .000$ (p = .05 divided by number of measures (k=114) across maths and geography, time 1, 2 & 3).

Differences between geography classes at time 3

School 1. ANOVA results by geography classroom at time 3 for school 1, without controlling for prior achievement, are presented in Table 5.1.15. The results show for the majority of measures, no significant differences between geography classrooms following multiple testing correction of $p \leq .000$ (p = .05/114). Geography performance (p = .001), student-teacher relations (p = .001), and geography anxiety (p = .003) fell just below significance. Levene's tests revealed that equal variances were assumed for most measures except classroom environment and classroom chaos (see Appendix 5, Table 5.1.3). The following two measures were significant:

Geography classroom environment time 3. Modest significant differences between classrooms were found for classroom environment, $F(7,167) = 4.331$, $p < .001$, $\eta_p^2 = .154$, with the highest mean score revealed for C6se and the lowest for C5se. Following multiple testing correction of $p \leq .000$ (p = .05/114), students, on average, did not rate their classroom environments significantly better than students in other classrooms. Levene's test showed unequal variances for these analyses (p = .024), with the least variance for C6se (0.35) and the most for C1e (1.14).

***Geography environment time 3*.** Moderate significant differences between classrooms were found for geography environment, $F(7,163) = 7.595$, $p < .001$, $\eta_p^2 = .246$, with the highest mean score revealed for C6se and the lowest for C8se. Pairwise comparisons showed that students in class C6se, on average, rated their geography environment in the use of equipment such as compasses etc., more highly than students in C4se, C7se and C8se ($p < .001$), following multiple testing correction of $p \leq .000$ (p = .05/114). Levene's test showed equal variances were assumed for these analyses (p = .339).

Table 5.1.15. Geography classroom variables at time 3 for school 1 (Russian sample): Means, standard deviation (SD), and N with ANOVA results by classroom without controlling for prior achievement

Construct	C1e	C2e	C3e	C4se	C5se	C6se	C7se	C8se	*p*	η_p^2
Geography performance	-0.18 (0.91) n=24	-0.07 (0.77) n=22	-0.13 (0.91) n=11	-0.05 (1.00) n=25	-0.33 (0.96) n=19	0.65 (0.77) n=29	0.54 (0.74) n=22	0.21 (1.00) n=26	.001	.134
Geography self-perceived ability	0.08 (0.95) n=24	-0.13 (0.81) n=22	0.19 (1.04) n=9	-0.29 (1.04) n=25	0.14 (1.06) n=20	-0.02 (0.96) n=28	0.47 (0.86) n=22	-0.07 (1.20) n=27	.324	.046
Geography enjoyment	0.30 (1.04) n=20	-0.04 (0.61) n=22	0.24 (0.66) n=9	-0.05 (0.89) n=25	-0.26 (1.25) n=20	-0.09 (0.88) n=27	0.28 (0.86) n=22	-0.37 (1.27) n=27	.220	.055
Geography classroom environment	-0.26 (1.00) n=24	0.21 (0.86) n=22	0.23 (0.86) n=9	0.10 (1.07) n=25	-0.37 (0.90) n=20	0.77 (0.59) n=28	0.28 (0.63) n=21	-0.10 (0.85) n=26	**.000**	**.154**
Geography classroom Student-teacher	-0.23 (0.95) n=23	0.33 (0.76) n=22	0.49 (0.81) n=9	0.14 (0.99) n=25	-0.42 (0.96) n=20	0.65 (0.74) n=28	0.14 (1.01) n=22	-0.25 (1.07) n=27	.001	.134
Geography classroom peer competition	-0.21 (1.12) n=23	0.09 (1.10) n=21	-0.14 (0.97) n=9	0.23 (1.12) n=25	-0.04 (0.80) n=20	0.39 (0.82) n=28	-0.06 (0.81) n=22	0.05 (1.05) n=27	.489	.037

Significant results in bold following a Bonferroni multiple testing correction of $p \leq .000$ ($p = .05$ divided by number of measures (k=114) across maths and geography, time 1, 2 & 3).

Table 5.1.15. Continued. Geography classroom variables at time 3 for school 1 (Russian sample): Means, standard deviation (SD), and N with ANOVA results by classroom without controlling for prior achievement

Construct	C1e	C2e	C3e	C4se	C5se	C6se	C7se	C8se	*p*	η_p^2
Geography classroom chaos	-0.11 (1.05) n=23	0.31 (0.97) n=22	0.06 (1.38) n=9	-0.15 (1.13) n=24	-0.68 (1.18) n=20	0.27 (0.78) n=27	0.31 (0.84) n=22	-0.23 (0.86) n=26	.017	.097
Geography homework behaviour	-0.16 (0.92) n=21	0.37 (0.88) n=21	-0.21 (1.00) n=9	0.01 (1.01) n=24	-0.04 (0.89) n=20	-0.05 (1.02) n=28	-0.32 (0.82) n=22	0.10 (1.14) n=27	.480	.038
Geography homework feedback	-0.23 (0.95) n=21	0.27 (0.75) n=21	0.12 (1.27) n=9	-0.11 (0.99) n=24	-0.33 (1.00) n=20	0.32 (0.85) n=28	0.24 (1.09) n=22	-0.24 (1.28) n=27	.177	.059
Geography homework total scale	-0.08) (0.96 n=21	0.04 (0.80) n=21	0.18 (1.10) n=9	-0.11 (0.80) n=24	-0.23 (0.78) n=20	0.26 (1.02) n=28	0.33 (1.04) n=22	-0.14 (1.19) n=26	.453	.040
Geography environment	-0.19 (0.87) n=20	0.21 (0.75) n=22	-0.03 (1.24) n=9	-0.25 (0.80) n=23	-0.25 (1.06) n=20	1.04 (0.86) n=28	-0.35 (0.81) n=22	-0.40 (0.87) n=27	**.000**	**.246**
Geography usefulness	-0.09 (0.95) n=22	-0.01 (0.93) n=22	0.06 (1.73) n=7	0.01 (0.80) n=23	-0.27 (1.00) n=20	0.17 (1.14) n=28	0.23 (1.18) n=22	-0.29 (1.13) n=26	.652	.030
Geography anxiety	-0.04 (0.87) n=23	0.18 (0.86) n=22	-0.23 (0.94) n=9	0.80 (1.15) n=25	-0.17 (0.98) n=19	-0.02 (0.97) n=28	-0.25 (0.71) n=22	-0.18 (0.96) n=27	.003	.118

Significant results in bold following a Bonferroni multiple testing correction of $p \leq .000$ ($p = .05$ divided by number of measures (k=114) across maths and geography, time 1, 2 & 3).

Table 5.1.16. Geography classroom variables at time 3 for school 2 (Russian sample): Means, standard deviation (SD), and N with ANOVA results by classroom without controlling for prior achievement

Construct	C9ce	C10ce	C11ce	*p*	η_p^2	Construct	C9ce	C10ce	C11ce	*p*	η_p^2
Geography performance	0.05 (0.99) n=16	0.11 (0.75) n=11	-1.38 (0.79) n=10	**.000**	**.375**						
Geography self-perceived ability	-0.03 (0.73) n=16	0.16 (0.75) n=11	0.03 (1.18) n=10	.846	.010	Geography homework behaviour	-0.32 (1.02) n=16	-0.15 (0.64) n=9	0.75 (0.89) n=10	.018	.223
Geography enjoyment	0.20 (0.78) n=16	0.17 (0.84) n=10	0.77 (1.17) n=9	.266	.079	Geography homework feedback	0.12 (0.89) n=16	0.12 (0.75) n=9	0.07 (1.22) n=9	.990	.001
Geography classroom environment	-0.46 (0.88) n=16	0.34 (0.87) n=11	-0.46 (0.79) n=10	.046	.166	Geography homework total scale	0.21 (1.04) n=16	0.26 (0.68) n=9	-0.26 (1.27) n=9	.471	.047
Geography classroom Student-teacher	-0.57 (0.84) n=16	0.21 (0.89) n=11	-0.43 (1.14) n=11	.110	.118	Geography environment	0.01 (0.94) n=16	0.73 (0.63) n=11	-0.40 (1.02) n=10	.018	.211
Geography classroom peer competition	-0.30 (0.93) n=16	0.11 (0.88) n=11	-0.90 (1.14) n=10	.075	.141	Geography usefulness	0.17 (0.81) n=16	0.00 (0.71) n=11	0.24 (0.68) n=10	.751	.017
Geography classroom chaos	0.17 (0.74) n=16	0.90 (0.46) n=11	-0.34 (0.86) n=11	.001	.325	Geography anxiety	-0.30 (0.83) n=16	-0.40 (0.52) n=11	-0.27 (0.68) n=10	.901	.006

Significant results in bold following a Bonferroni multiple testing correction of $p \leq .000$ (p = .05 divided by number of measures (k=114) across maths and geography, time 1, 2 & 3).

School 2. ANOVA results by geography classroom at time 3 for school 2, without controlling for prior achievement, are presented in Table 5.1.16. The results show for the majority of measures, no significant differences between geography classrooms following multiple testing correction of $p \leq .000$ (p = .05/114). Geography classroom chaos (p = .001) fell just below significance. Levene's tests revealed that equal variances were assumed for all measures (see Appendix 5, Table 5.1.4). The following measure was significant:

***Geography performance time 3*.** Moderate significant differences between classrooms were found for geography performance, $F(2,34) = 10.198$, $p < .001$, $\eta_p^2 = .375$, with the highest mean score revealed for C10ce and the lowest for C11ce. Following multiple testing correction of $p \leq .000$ (p = .05/114), no pairwise comparisons reached significance, the difference between class C11ce with the lowest performance, and the other two classes fell just below significance ($p \leq .001$). Levene's test showed equal variances were assumed for these analyses (p = .251).

Maths and geography teacher group differences at time 3

ANOVA results at time 3, without controlling for prior achievement, can be found for maths teachers in Table 5.1.17, and for geography teachers in Table 5.1.18. The results show for the majority of the measures, no significant differences between maths or geography teacher groupings following multiple testing correction of $p \leq .001$ (p = .05/57). Measures just below the significance threshold were: maths performance (p = .002), maths environment (p =.009), and maths homework behaviour (p = .003). Levene's tests revealed that equal variances were assumed for most measures, except maths classroom environment, maths student-teacher relations, maths peer competition, geography enjoyment, geography classroom environment, geography peer competition, geography classroom chaos and geography anxiety (see Appendix 5, Tables 5.1.5 and 5.1.6). The following seven measures were significant:

Maths classroom environment time 3. Modest significant differences between teacher groupings were found for maths classroom environment, $F(5,205) = 5.577$, $p < .001$, $\eta_p^2 = .120$, with the highest mean score revealed for TM4 and the lowest for TM1. Following multiple testing correction of $p \leq .001$ ($p = .05/57$), students on average, did not rate their classroom environment differently, when taught by different teachers. Levene's test showed unequal variances for these analyses ($p = .004$) with the smallest variance shown for teacher TM4 (0.48) who teaches just one class and the largest for teacher TM2 (1.59) who teaches two classes.

***Maths student-teacher relations time 3*.** Modest significant differences between teacher groupings were found for student-teacher relations, $F(5,205) = 6.363$, $p < .001$, $\eta_p^2 = .134$, with the highest mean score revealed for TM4 and the lowest for TM1. Following multiple testing correction of $p \leq .001$ ($p = .05/57$), students on average, did not rate their student-teacher relations differently as a result of being taught by different maths teachers. Levene's test showed unequal variances for these analyses ($p = .004$) with the smallest variance shown for teacher TM6 (0.55) who teaches four classes the largest shown for teacher TM2 who teaches two classes (1.35).

Table 5.1.17. Maths teacher groups time 3 (Russian sample): Means, standard deviation (SD) and N for maths classroom variables with ANOVA results by teacher group, without controlling for prior achievement

Construct	TM1	TM2	TM3	TM4	TM5	TM6	*p*	η_p^2
Maths performance	0.38 (0.91) n=16	0.14 (1.07) n=22	0.20 (0.84 n=41	-0.22 (0.95) n=11	0.53 (0.86) n=29	-0.24 (1.01) n=92	.002	.088
Number line	-0.55 (1.14) n=15	0.40 (1.00) n=22	0.19 (0.86) n=46	-0.02 (0.92) n=11	-0.16 (0.89) n=29	0.19 (0.93) n=91	.024	.060
Maths self-perceived ability	-0.34 (0.86) n=16	0.39 (1.04) n=19	-0.07 (0.93 n=45	0.09 (0.90) n=11	0.06 (1.06) n=28	0.01 (1.02) n=93	.407	.024
Maths enjoyment	-0.08 (1.13) n=16	0.17 (0.92) n=18	0.06 (0.89 n=41	0.33 (0.85) n=11	0.09 (1.24) n=27	-0.06 (0.98) n=91	.788	.012
Maths classroom environment	-0.39 (0.93) n=16	-0.33 (1.26) n=19	0.30 (0.73 n=45	0.66 (0.69) n=11	0.47 (0.70) n=29	-0.04 (0.76) n=91	**.000**	**.120**
Maths classroom student-teacher relations	-0.42 (0.98) n=16	-0.27 (1.16) n=19	0.35 (0.75 n=45	0.68 (0.79) n=11	0.50 (0.81) n=29	-0.09 (0.74) n=91	**.000**	**.134**
Maths classroom peer competition	-0.25 (1.00) n=16	-0.56 (1.34) n=21	-0.04 (0.90 n=45	0.35 (0.58) n=11	0.29 (0.74) n=29	0.12 (0.95) n=93	.019	.062

Significant results in bold following a Bonferroni multiple testing correction of $p \leq .001$ ($p = .05$ divided by number of measures (k=57) across maths and geography, time 1, 2 & 3).

Table 5.1.17. Continued. Maths teacher groups time 3 (Russian sample): Means, standard deviation (SD) and N for maths classroom variables with ANOVA results by teacher group, without controlling for prior achievement

Construct	TM1	TM2	TM3	TM4	TM5	TM6	*p*	η_p^2
Maths classroom chaos	0.02 (0.87) n=16	0.17 (0.84) n=22	0.14 (1.08) n=46	0.10 (1.320 n=11	0.24 (0.98) n=29	-0.12 (0.96) n=94	.460	.022
Maths homework behaviour	-0.10 (0.87) n=16	0.75 (0.80) n=22	0.06 (0.90) n=44	0.51 (0.51) n=11	-0.35 (0.97) n=29	0.01 (1.09) n=93	.003	.083
Maths homework feedback	0.03 (0.61) n=16	0.20 (1.15) n=22	0.24 (0.90) n=44	-0.30 (0.90) n=11	0.10 (1.18) n=29	-0.13 (1.01) n=92	.288	.029
Maths homework total scale	0.16 (0.69) n=16	-0.22 (1.02) n=22	0.22 (0.72) n=44	-0.35 (0.76) n=11	0.26 (1.15) n=29	-0.02 (0.98) n=91	.190	.035
Maths environment	0.37 (0.96) n=15	-0.12 (1.09) n=21	0.31 (0.91) n=45	0.05 (1.09) n=11	0.40 (0.91) n=28	-0.21 (0.95) n=91	.009	.072
Maths usefulness	0.05 (0.91) n=16	0.62 (0.92) n=22	-0.22 (0.75) n=44	-0.01 (0.94) n=9	-0.17 (0.92) n=28	-0.10 (1.01) n=93	.018	.064
Maths anxiety	-0.15 (0.95) n=15	-0.54 (0.85) n=21	0.12 (0.86) n=44	-0.30 (0.77) n=11	-0.16 (0.98) n=27	0.09 (1.03) n=93	.080	.047

Significant results in bold following a Bonferroni multiple testing correction of $p \leq .001$ ($p = .05$ divided by number of measures (k=57) across maths and geography, time 1, 2 & 3).

Table 5.1.18. Geography teacher groups time 3 (Russian sample): Means, standard deviation (SD) and N for geography classroom variables with ANOVA results by teacher group, without controlling for prior achievement

Construct	TG1	TG2	TG3	TG4	TG5	*p*	η_p^2
Geography performance	-0.32 (1.07) n=37	0.17 (0.96) n=67	0.65 (0.77) n=29	-0.13 (0.84) n=46	-0.07 (0.96) n=36	**.000**	**.092**
Geography self-perceived ability	0.04 (0.86) n=37	0.16 (1.07) n=69	-0.02 (0.96) n=28	-0.02 (0.89) n=46	-0.16 (1.05) n=34	.608	.013
Geography enjoyment	0.34 (0.92) n=35	-0.13 (1.17) n=69	-0.09 (0.88) n=27	0.12 (0.85) n=42	0.03 (0.83) n=34	.196	.029
Geography classroom environment	-0.22 (0.91) n=37	-0.06 (0.83) n=67	0.77 (0.59) n=28	-0.03 (0.95) n=46	0.13 (1.01) n=34	**.000**	**.105**
Geography classroom Student-teacher relations	-0.30 (0.99) n=38	-0.17 (1.03) n=69	0.65 (0.74) n=28	0.04 (0.90) n=45	0.23 (0.94) n=34	**.000**	**.093**
Geography classroom peer competition	-0.34 (1.02) n=37	-0.01 (0.90) n=69	0.39 (0.82) n=28	-0.07 (1.11) n=44	0.13 (1.08) n=34	.052	.044

Significant results in bold following a Bonferroni multiple testing correction of $p \leq .001$ ($p = .05$ divided by number of measures (k=57) across maths and geography, time 1, 2 & 3).

Table 5.1.18. Continued. Geography teacher groups time 3 (Russian sample): Means, standard deviation (SD) and N for geography classroom variables with ANOVA results by teacher group, without controlling for prior achievement

Construct	TG1	TG2	TG3	TG4	TG5	*p*	η_p^2
Geography classroom chaos	0.24 (0.84) n=38	-0.19 (1.02) n=68	0.27 (0.78) n=27	0.10 (1.02) n=45	-0.09 (1.19) n=33	.133	.033
Geography homework behaviour	0.03 (0.99) n=35	-0.07 (0.98) n=69	-0.05 (1.02) n=28	0.10 (0.93) n=42	-0.05 (0.99) n=33	.910	.005
Geography homework feedback	0.10 (0.93) n=34	-0.11 (1.16) n=69	0.32 (0.85) n=28	0.02 (0.88) n=42	-0.05 (1.06) n=33	.414	.019
Geography homework total scale	0.10 (1.02) n=34	-0.01 (1.05) n=68	0.26 (1.02) n=28	-0.02 (0.88) n=42	-0.03 (0.88) n=33	.735	.010
Geography environment	0.11 (0.97) n=37	-0.34 (0.90) n=69	1.04 (0.86) n=28	0.02 (0.82) n=42	-0.19 (0.93) n=32	**.000**	**.196**
Geography usefulness	0.14 (0.74) n=37	-0.11 (1.12) n=68	0.17 (1.14) n=28	-0.05 (0.93) n=44	0.02 (1.05) n=30	.649	.012
Geography anxiety	-0.32 (0.69) n=37	-0.20 (0.88) n=68	-0.02 (0.97) n=28	0.07 (0.86) n=45	0.53 (1.18) n=34	**.001**	**.085**

Significant results in bold following a Bonferroni multiple testing correction of $p \leq .001$(p = .05 divided by number of measures (k=57) across maths and geography, time 1, 2 & 3).

***Geography performance time 3*.** Small significant differences between teacher groupings were found for geography performance, $F(4,210) = 5.321$, $p < .001$, $\eta_p^2 = .092$, with the highest mean score revealed for TG3 and the lowest for TG1. Following multiple testing correction of $p \leq .001$ ($p = .05/57$), pairwise comparisons showed that students studying geography with teacher TG3, on average, had performed significantly better on the task than students studying geography with teachers TG1 and TG4 ($p < .001$). Levene's test showed equal variances were assumed for these analyses ($p = .259$).

***Geography classroom environment time 3*.** Modest significant differences between teacher groupings were found for classroom environment, $F(4,207) = 6.041$, $p < .001$, $\eta_p^2 = .105$, with the highest mean score revealed for TG3 and the lowest for TG1. Following multiple testing correction of $p \leq .001$ ($p = .05/57$), pairwise comparisons showed that students studying geography with teacher TG3, on average, rated their classroom environment significantly better than students studying geography with teachers TG1, TG2 and TG4 ($p < .001$). Levene's test showed unequal variances for these analyses ($p = .044$), with the smallest variance revealed for teacher TG3 (0.35) who teaches one class and the largest for TG5 (1.02) who teaches two classes..

***Geography student-teacher relations time 3*.** Small significant differences between teacher groupings were found for student-teacher relations, $F(4,209) = 5.340$, $p < .001$, $\eta_p^2 = .093$, with the highest mean score revealed for TG3 and the lowest for TG1. Following multiple testing correction of $p \leq .001$ ($p = .05/57$), pairwise comparisons showed that students studying geography with teacher TG3, on average, rated their student-teacher relations significantly higher than students studying geography with teachers TG1 and TG2 ($p < .001$). Levene's test showed equal variances were assumed for these analyses (p = .454).

***Geography environment time 3*.** Modest significant differences between teacher groupings were found for geography environment, $F(4,203) = 12.349$, $p < .001$, $\eta_p^2 = .196$, with the highest mean score revealed for TG3 and the lowest for TG2. Following multiple testing correction of $p \leq .001$ (p = .05/57), pairwise comparisons showed that students studying geography with teacher TG3, on average, rated their geography environment in the use of equipment such as compasses etc., more highly than students studying geography with TG2, TG4 and TG5 ($p < .001$). Levene's test showed equal variances were assumed for these analyses (p = .651).

***Geography anxiety time 3*.** Small significant differences between teacher groupings were found, $F(4,207) = 4.815$, $p = .001$, $\eta_p^2 = .085$, with the highest mean score revealed for TG5 and the lowest for TG1 (high score indicates high anxiety). Following multiple testing correction of $p \leq .001$ (p = .05/57), students on average, did not perceive levels of geography anxiety differently as a result of being taught by a different teacher. Levene's test showed unequal variances for these analyses (p = .003) with the smallest variance shown for teacher TG1 (0.48) who teaches three classes and the largest shown for teacher TG5 (1.39) who teaches two classes.

Class and teacher group ranking by mean score at time 3

Maths classroom. The results in Table 5.1.19 for school 1,without controlling for prior achievement, show some variation of ranking position for most classrooms across the measures. Classes C4se, C5se, and C6se show the most correspondence, the remaining classes are more mixed. Classes C4se, C5se, C7se and C8se have lower placed ranks. Class C2e ranks highest across the measures in first, second and third place. A modest amount of correspondence of ranking is shown with maths classes ranked at time 1 (Chapter 4, p 139).

The results for school 2, without controlling for prior achievement, in Table 5.1.20 with fewer classes, show more consistency in rank across the measures. Class 10ce is ranked highest for four out of five measures and C11ce is ranked in third (lowest) place.

Geography classroom time 3. The results for school 1, without controlling for prior achievement, in Table 5.1.21 show a similar level of correspondence of rank across the measures as for maths classroom. Classes C1e, C5se, and C8se sit at the lower ranks, C6se sits in first place for four out of five measures, the other classrooms show more variation across the measures. Modest correspondence is shown with time 1 geography classroom rankings with C6se predominantly in first place at both assessment waves (Chapter 4, p. 141).

The results for school 2, without controlling for prior achievement, in Table 5.1.22 show complete correspondence of rank across the two measures. C9ce is in first place, and C11ce is in third.

Maths teacher time 3. The results without controlling for prior achievement, in Table 5.1.23 show some consistency of rank across the measures for most teacher groups. TM3 is consistently in third place across the four measures. TM2 ranks in fifth place for three out of four measures. All other groups rank in no more than two places across the measures. Somewhat similar levels of correspondence are shared with time 1 (Chapter 4, p. 154).

Table 5.1.19. Maths classroom variables at time 3 for school 1 (Russian sample): Classrooms ranked by means (highest = 1 to lowest = 8) for measures demonstrating a significant effect of maths classroom, without controlling for prior achievement

Construct	C1e	C2e	C3e	C4se	C5se	C6se	C7se	C8se	*p*	η_p^2
Maths performance	2nd	3rd	6th	5th	8th	1st	7th	4th	.002	.127
Maths classroom environment	4th	2nd	1st	5th	8th	3rd	7th	6th	**.000**	**.142**
Maths classroom Student-teacher relations	5th	2nd	1st	7th	8th	3rd	6th	4th	**.000**	**.187**
Maths classroom chaos	7th	1st	4th	5th	6th	3rd	2nd	8th	.001	.134

Significant results in bold following a Bonferroni multiple testing correction of $p \leq .000$ ($p = .05$ divided by number of measures (k=114) across maths and geography, time 1, 2 & 3).

Table 5.1.20. Maths classroom variables at time 3 for school 2 (Russian sample): Classrooms ranked by means (highest = 1 to lowest = 3) for measures demonstrating a significant effect of maths classroom, without controlling for prior achievement

Construct	C9ce	C10ce	C11ce	p	η_p^2
Maths classroom environment	**2nd**	**1st**	**3rd**	**.000**	**.495**
Maths classroom environment	**2nd**	**1st**	**3rd**	**.000**	**.495**
Maths classroom student-teacher relations*	**2nd**	**1st**	**3rd**	.001	.373
Maths classroom peer competition*	**2nd**	**1st**	**3rd**	.001	.339
Maths homework behaviour	**1st**	**2nd**	**3rd**	**.000**	**.395**

Significant results in bold following a Bonferroni multiple testing correction of $p \leq .000$ ($p = .05$ divided by number of measures (k=114) across maths and geography, time 1, 2 & 3). *Just below significance, ranked for parity.

Table 5.1.21. Geography classroom variables at time 3 for school 1 (Russian sample): Classrooms ranked by means (highest = 1 to lowest = 8) for measures demonstrating a significant effect of geography classroom at time 2, without controlling for prior achievement

Construct	C1e	C2e	C3e	C4se	C5se	C6se	C7se	C8se	*p*	η_p^2
Geography performance*	7th	5th	6th	4th	8th	1st	2nd	3rd	.001	.134
Geography classroom environment	7th	4th	3rd	5th	8th	1st	2nd	6th	**.000**	**.154**
Geography classroom Student-teacher*	6th	3rd	2nd	4th	8th	1st	5th	7th	.001	.134
Geography environment	4th	2nd	3rd	5th	6th	1st	7th	8th	**.000**	**.246**
Geography anxiety*	4th	2nd	7th	1st	5th	3rd	8th	6th	.003	.118

Significant results in bold following a Bonferroni multiple testing correction of $p \leq .000$ (p = .05 divided by number of measures (k=114) across maths and geography, time 1, 2 & 3) . *Just below significance, ranked for parity.

Table 5.1.22. Geography classroom variables at time 3 for school 2 (Russian sample): Classrooms ranked by means (highest = 1 to lowest = 3) for measures demonstrating a significant effect of geography classroom, without controlling for prior achievement

Construct	C9ce	C10ce	C11ce	*p*	η_p^2
Geography performance	**1st**	**2nd**	**3rd**	**.000**	**.375**
Geography classroom chaos*	**1st**	**2nd**	**3rd**	.001	.325

Significant results in bold following a Bonferroni multiple testing correction of $p \leq .000$ ($p = .05$ divided by number of measures (k=114) across maths and geography, time 1, 2 & 3). *Just below significance, ranked for parity.

Table 5.1.23. Maths Teacher groups at time 3 (Russian sample): ranked by means (highest = 1 to lowest = 6) for measures demonstrating a significant effect of maths teacher without controlling for prior achievement

Construct	TM1	TM2	TM3	TM4	TM5	TM6	*p*	η_p^2
Maths performance*	**2nd**	**4th**	**3rd**	**5th**	**1st**	**6th**	.002	.088
Maths classroom environment	**6th**	**5th**	**3rd**	**1st**	**2nd**	**4th**	**.000**	**.120**
Maths classroom student-teacher relations	**6th**	**5th**	**3rd**	**1st**	**2nd**	**4th**	**.000**	**.134**
Maths environment*	**2nd**	**5th**	**3rd**	**4th**	**1st**	**6th**	.009	.072

Significant results in bold following a Bonferroni multiple testing correction of $p \leq .001$ ($p = .05$ divided by number of measures (k=57) across maths and geography, time 1, 2 & 3). *Just below significance, ranked for parity.

Table 5.1.24. Geography Teacher groups at time 3 (Russian sample): ranked by means (highest = 1 to lowest = 5) for measures demonstrating a significant effect of geography teacher without controlling for prior achievement

Construct	TG1	TG2	TG3	TG4	TG5	*p*	η_p^2
Geography performance	**5th**	**2nd**	**1st**	**4th**	**3rd**	**.000**	**.092**
Geography classroom environment	**5th**	**4th**	**1st**	**3rd**	**2nd**	**.000**	**.105**
Geography classroom Student-teacher relations	**5th**	**4th**	**1st**	**3rd**	**2nd**	**.000**	**.093**
Geography environment	**2nd**	**5th**	**1st**	**3rd**	**4th**	**.000**	**.196**
Geography anxiety	**5th**	**4th**	**3rd**	**2nd**	**1st**	**.001**	**.085**

Significant results in bold following a Bonferroni multiple testing correction of $p \leq .001$ ($p = .05$ divided by number of measures (k=57) across maths and geography, time 1, 2 & 3). *Just below significance, ranked for parity.

Geography teacher. The results without controlling for prior achievement, in Table 5.1.24 show slightly more consistency of rank across measures than maths teacher groups. The most consistent groups are TG1, ranking fifth place and TG3 ranking first, four times out of five. TG2 ranks fourth, and TG4 ranks third, three times out of five. TG5 shows the most variation, ranking in four positions across the measures. Very little correspondence is shown with time 1 apart from teacher TG3 who ranks predominantly in first place at both assessment waves (Chapter 4, p. 155).

5.1 Discussion

The aim of part 5.1 was to investigate the research question of whether significant effects of classroom and teacher groups found at time 1 persisted across time 2 and time 3, without controlling for primary school achievement. The significant differences between classrooms for some measures at time 2 with modest effect sizes ranging from 14.1% to 21.3% (see Tables 5.1.1 to 5.1.4) suggest some similarity with results at time 1. However, slightly fewer measures reached significance following multiple testing correction. For maths classroom, differences were shown for different measures than at time 1. For example, at time 2 significant differences between maths classrooms were observed for the number line task and homework feedback rather than classroom environment and student-teacher relations. These findings suggest more variation between classrooms in number estimation and homework feedback, but greater similarity in terms of classroom environment and student-teacher relations than at time 1. This is an interesting finding given the mixed ability classrooms and standardised curricula. More divergence would be expected on average across classrooms for measures such as student-teacher relations rather than mathematical ability by this stage of the academic year. For geography, significant differences between classrooms were observed for largely the same measures as at time 1, albeit fewer. Similar findings were also observed for teacher groups but with smaller effects than observed at time 1, ranging from 9.6% to 11.8% for maths teacher groups (see Table 5.1.5). Similarly to maths classrooms, differences between maths teacher groups were shown for some additional measures to time 1. The number line task and peer competition were now significant, student-teacher relations fell just below significance. For geography time 2, significant differences between teacher groups were observed for only one measure, geography environment, with a moderate effect (18%) (see Table 5.1.6). Together the findings suggest a weaker effect of classroom and teacher groups at time 2.

At time 3, the effect of classroom and teacher groups appeared to weaken further as even fewer significant differences were observed across maths and geography classrooms and teacher groups. Only maths classroom and student-teacher relations revealed significant differences between classrooms, although effect sizes were modest (14.2% and 18.7%, respectively) (see Table 5.1.13). Equally for geography classrooms, significant differences were only revealed for student-teacher relations and geography environment, albeit with slightly stronger effect sizes (15% and 24.6% respectively) (see Table 5.1.15). Similarly, only classroom environment and student-teacher relations showed significant differences between maths teacher groups, with modest effect sizes (12% and 13.4% respectively) (see Table 5.1.17). Unlike time 2, these measures reaching significance reflected the measures reaching significance at time 1. Only slightly more measures reached significance for geography teacher groups, however, with the interesting addition of geography anxiety showing small significant differences between teacher groups. Effect sizes were modest, ranging from 8.5% to 19.6% (see Table 5.1.18). The measure was just below the significance threshold for classrooms. The effect sizes overall for class and teacher groups were somewhat reduced to those found at time 1.

Part 5.1 also explored the research question of whether the patterns of class rankings of mean scores from highest to lowest, found at time 1 were also maintained across subsequent waves. If the influence of class or teacher group is strong, then more consistency of ranking position would be expected across the measures for class and teacher groups. The larger amount of variation across measures at time 2 (Tables 5.1.7 to 5.1.12) compared to time 1 (Chapter 4 p. 139 to 155) however, suggests a weakening effect of maths classroom. A slightly stronger influence of geography classroom is evident however, as more consistency was observed across the measures than for maths classroom. Similarly to maths classroom, however, there is less agreement with the ranking patterns found at time 1 (Chapter 4 p. 141). The ranking patterns for the teacher groups show slightly less variation across the measures than seen for classrooms. Some correspondence is shown with time 1, especially with classroom environment and student-teacher relations for maths teacher groups. An interesting finding, which suggests some effect of teacher group across the two waves. However, it is likely that students also contribute towards this effect, rather than a dominant effect of teacher.

The ranking patterns at time 3 also show some variation across the measures for both maths and geography classrooms (Table 5.1.19 and 5.1.21). There are some classes that show more correspondence. For example, class C6se ranks mainly in first place across time 2 and time 3 for geography classrooms. Geography teacher groups show slightly more consistency than maths teacher groups (see Tables 5.1.23 and 5.1.24). However, very little correspondence is shown with time 1, apart from teacher TG3, who only teaches class C6se and ranks predominantly in first place across time 1, 2 and 3.

The findings in part 5.1 suggest some similarity with findings at time 1 with modest effects observed for a few measures of maths and geography classroom at time 2. However, fewer measures reached significance at time 3 suggesting that any effect at time 1 was weakening by time 3. Greater variation in ranking for time 2 and 3 than at time 1, and less agreement with rank positions at time 1 suggests more divergence from time 1 effects. The consistency observed for specific classrooms indicates that it may be specific classes that are strengthening any effect. Several of the measures showing significant effects were also subject to unequal variances, preventing complete confidence in interpreting the results. For classrooms no particular pattern was observed of more variance for certain classes, suggesting a degree of randomness across measures showing unequal variance. For teacher groups, the pattern was more consistent with teachers TM2 and TG5 demonstrating more variance. It might be expected that teachers with more classes would demonstrate more variance, however, this does not appear to be so as although both TM2 and TG5 teach two classes, other teachers cover as many as four classes, without showing any difference in variance. Overall the findings suggest a weakening effect of classroom/teacher groups observed at time 1 for measures of maths and geography classrooms, without controlling for primary school achievement.

5.2. Classroom And Teacher Differences At Time 2 And Time 3 In The Russian Sample, Controlling For Prior Achievement

To establish whether or not patterns in results found at time 2 and time 3 were largely due to the primary school teacher or prior selection processes, the analyses were repeated while controlling for primary school achievement. Maths and geography grades were collected from students' final year in primary education. The maths and geography study measures and year 5 school achievement were regressed on students' maths and geography grades, respectively. ANOVAs were conducted using these new variables. To provide a more direct comparison with time 1, analyses for maths performance at time 1 were also conducted by classroom and by teacher while controlling for prior achievement.

Means, standard deviations (SD) and N for all assessed variables by classroom and by teacher, are presented in Tables 5.2.1 to 5.2.3 for time 1; and in Appendix 5, Tables 5.2.5 to 5.2.11 for time 2; and Tables 5.2.17 to 5.2.22 for time 3.

Differences for maths performance at time 1 by classroom

ANOVA results for school 1 and school 2 can be seen in Tables 5.2.1 and 5.2.2, Levene's test results are presented in Appendix 5, Tables 5.2.29 and 5.2.30. A Bonferroni multiple testing correction was set of $p \leq .000$ where $p = .05$ divided by the number of measures (k=114) across both schools and maths and geography at time 1, time 2 and time 3.

Moderate significant differences between classrooms were observed for school 1 only in maths performance at time 1, $F(7,164) = 7.537$, $p < .001$, $\eta_p^2 = .243$, with the highest mean score shown for C6se and the lowest for C5se. Pairwise comparisons revealed that students in class C6se, on average performed significantly better than students in class C5se ($p < .001$), following a multiple testing correction of $p \leq .000$($p = .05 /114$). Levene's test showed equal variances were assumed for these analyses ($p = .586$). These results show a slightly reduced effect of classroom and fewer significant pairwise comparisons when controlling for prior achievement, compared to the previous analysis in Chapter 4.

Differences for maths performance at time 1 by teacher group

ANOVA results by teacher can be seen in Table 5.2.3 A Bonferroni multiple testing correction was set of $p \leq .001$ where $p = .05$ divided by the number of measures (k=57) across maths and geography at time 1, time 2 and time 3.

Moderate significant differences between teacher groupings were revealed for maths performance at time 1, $F(5,201) = 12.010$, $p < .001$, $\eta_p^2 = .230$, with the highest mean score for TM1 and the lowest for TM6. Following a multiple testing correction of $p \leq .000$ ($p = .05 /114$), pairwise comparisons showed that students studying maths with teachers TM1 ($p < .001$), TM2 ($p < .001$), and TM5 ($p < .001$), on average, performed significantly better than students studying with teacher TM6. Levene's test showed equal variances were assumed for these analyses (*p* = .185) (see Appendix 5, Table 5.2.33). In contrast to the classroom analysis above, these results show that when controlling for prior achievement the effect of teacher is slightly increased and there is no reduction in the number of pairwise comparisons compared to the analysis in Chapter 4. In this analysis, a different teacher group has the highest mean score, whereas TM5 was the highest previously.

Class and teacher group ranking for maths performance at time 1

Maths classroom. The classes were ranked by their mean scores, from highest to lowest, to assess ranking positions for maths performance by classroom with prior achievement controlled for. Table 5.2.4 shows the results in comparison with the analysis in Chapter 4, where prior achievement was not controlled. We can see no change in rank for C6se, C2eand C5se, in first, second and eighth place, respectively. For the other classes, some changed slightly, up or down a rank (C1e, C4se, and C7se), but C3e and C8se changed considerably.

Maths teacher groups. Teacher groups were also ranked by their mean scores, from highest to lowest, to compare ranking, with and without controlling for prior achievement. Table 5.2.5 shows no change in ranking for all teacher groups apart from TM1, which changes to first place, and TM5, which changes to second place.

Table 5.2.1. Maths performance at time 1 for school 1 (Russian sample): Means, standard deviation (SD) and N with ANOVA results by classroom with and without controlling for prior achievement

Construct	C1e	C2e	C3e	C4se	C5se	C6se	C7se	C8se	*p*	η_p^2
Maths performance prior achievement controlled for	0.22 (1.08) n=21	0.48 (0.93) n=9	-0.08 (1.05) n=18	-0.22 (0.97) n=27	-1.06 (0.81) n=23	0.59 (0.84) n=28	-0.24 (0.75) n=23	-0.40 (0.81) n=23	**.000**	**.243**
Maths performance prior achievement not controlled for	-0.08 (1.17) n=23	0.40 (1.17) n=9	-0.34 (1.03) n=18	-0.27 (0.86) n=28	-1.15 (0.66) n=25	0.69 (0.80) n=28	-0.21 (0.68) n=24	0.01 (0.81) n=31	**.000**	**.265**

Significant results in bold following a Bonferroni multiple testing correction of $p \leq .00$ ($p = .05$ divided by number of measures (k=114) across maths and geography, time 1, 2 & 3).

Table 5.2.2. Maths performance at time 1 for school 2 (Russian sample): Means, standard deviation (SD), and N with ANOVA results by classroom with and without controlling for prior achievement

Construct	C9ce	C10ce	C11ce	*p*	η_p^2
Maths performance prior achievement controlled for	0.72 (0.62) n=16	0.53 (0.78) n=10	0.48 (0.49) n=9	.623	.029
Maths performance prior achievement not controlled for	0.66 (0.83) n=18	0.56 (0.72) n=11	0.59 (0.56) n=14	.921	.004

Significant results in bold following a Bonferroni multiple testing correction of $p \leq .00$
($p = .05$ divided by number of measures (k=114) across maths and geography, time 1, 2 & 3).

Table 5.2.3. Maths performance at time 1 for maths teacher groups (Russian sample): Means, standard deviation (SD), and N with ANOVA results by classroom with and without controlling for prior achievement

Construct	TM1	TM2	TM3	TM4	TM5	TM6	*p*	η_p^2
Maths performance prior achievement controlled for	0.72 (0.62) n=16	0.51 (0.64) n=19	0.30 (1.03) n=30	-0.08 (1.05) n=18	0.59 (0.84) n=28	-0.47 (0.90) n=96	**.000**	**.230**
Maths performance prior achievement not controlled for	0.66 (0.83) n=18	0.57 (0.62) n=25	0.06 (1.17) n=32	-0.34 (1.03) n=18	0.69 (0.80) n=28	-0.38 (0.87) n=108	**.000**	**.208**

Significant results in bold following a Bonferroni multiple testing correction of $p \leq .001$($p = .05$ divided by number of measures (k=57) across maths and geography, time 1, 2 & 3).

Table 5.2.4. Maths performance at time 1 for school 1 (Russian sample): Classrooms ranked by means (highest =1 to lowest = 8) with and without controlling for prior achievement

Construct	C1e	C2e	C3e	C4se	C5se	C6se	C7se	C8se	*p*	η_p^2
Maths performance prior achievement controlled for	**3rd**	**2nd**	**4th**	**5th**	**8th**	**1st**	**6th**	**7th**	**.000**	**.243**
Maths performance prior achievement not controlled for	**4th**	**2nd**	**7th**	**6th**	**8th**	**1st**	**5th**	**3rd**	**.000**	**.265**

Significant results in bold following a Bonferroni multiple testing correction of $p \leq .000$ ($p = .05$ divided by number of measures (k=114) across maths and geography, time 1, 2 & 3.

Table 5.2.5. Maths performance at time 1 for maths teacher groups (Russian sample): Classrooms ranked by means (highest =1 to lowest = 6) with and without controlling for prior achievement

Construct	TM1	TM2	TM3	TM4	TM5	TM6	*p*	η_p^2
Maths performance prior achievement controlled for	1st	**3rd**	**4th**	**5th**	**2nd**	**6th**	**.000**	**.230**
Maths performance prior achievement not controlled for	**2nd**	**3rd**	**4th**	**5th**	**1st**	**6th**	**.000**	**.208**

Significant results in bold following a Bonferroni multiple testing correction of $p \leq .001$($p = .05$ divided by number of measures (k=57) across maths and geography, time 1, 2 & 3.

Differences Between Maths Classrooms At Time 2

School 1. ANOVA results by maths classroom at time 2, controlling for prior achievement, are presented in Appendix 5, Table 5.2.6. The results show that for the majority of measures, there were no significant differences between maths classrooms following multiple testing correction of $p \leq .000$($p = .05/114$). Maths performance fell just below the significance threshold ($p = .001$). Levene's tests revealed that equal variances were assumed for most measures except classroom chaos and maths environment, (see Appendix 5, Table 5.2.29). A significant effect of classroom was found for the following five measures:

***Maths Year 5 school achievement time 2*.** Modest significant differences between classrooms were found for year 5 school achievement, $F(7,157) = 3.998$, $p < .001$, $\eta_p^2 = .151$, with the highest mean score revealed for class C6se and the lowest for C8se. Following multiple testing correction of $p \leq .000$ ($p = .05/114$), pairwise comparisons showed that students on average, did not differ in their school achievement as a consequence of being in a specific class. Levene's test showed equal variances were assumed for these analyses ($p = .199$)

Number line time 2. Modest significant differences between classrooms were found for the number line task, $F(7,157) = 5.271$, $p < .001$, $\eta_p^2 = .190$, with the lowest mean score revealed for class C4se and the highest for C8se (an optimum score is low). Following multiple testing correction of $p \leq .000$ ($p = .05/114$), pairwise comparisons indicated that students in class C4se performed significantly better than class C8se ($p < .000$). Levene's test showed equal variances were assumed for these analyses ($p = .272$).

Maths classroom chaos time 2. Modest significant differences between classrooms were found for classroom chaos, $F(7,159) = 5.469$, $p < .001$, $\eta_p^2 = .194$, with the highest mean score revealed for class C7se and the lowest for C4se (a high score indicates low chaos). Following multiple testing correction of $p \leq .000$ ($p = .05/114$), pairwise comparisons revealed that students did not on average, differ in their ratings of classroom chaos as a result of being in a specific class. However, Levene's test showed unequal variances for these analyses ($p = .010$), with Class C2e having the smallest variance (0.45) and class C3e having the largest (1.70).

***Maths homework feedback time 2*.** Modest significant differences between classrooms were found for homework feedback, $F(7,156) = 4.041$, $p < .001$, $\eta_p^2 = .153$, with the highest mean score revealed for class C7se and the lowest for C1e. Following multiple testing correction of $p \leq .000$ ($p = .05/114$), pairwise comparisons revealed that students on average did not rate their homework feedback differently as result of being in a specific class. Levene's test showed unequal variances for these analyses ($p = .041$) with the smallest variance revealed for class C2e (0.31) and the largest for class C5se (1.42).

Maths environment time 2. Modest significant differences between classrooms were found for maths environment, $F(7,156) = 4.430$, $p < .001$, $\eta_p^2 = .166$, with the highest mean score revealed for class C6se and the lowest for C8se. Following multiple testing correction of $p \leq .000$ ($p = .05/114$) pairwise comparisons showed that students did not, on average, rate their maths environment differently in relation to the use of rulers and calculators etc., as a result of being in a specific class. Levene's test showed unequal variances for these analyses ($p = .045$).

School 2. ANOVA results for school 2 by maths classroom, controlling for prior achievement, are presented in Appendix 5, Table 5.2.7 and show no significant differences between maths classroom for all of the measures following multiple testing correction of $p \leq .000$ ($p = .05/114$). Maths Year 5 school achievement and maths classroom environment fell just below the significance threshold ($p = .009$ and $p = .001$, respectively). Levene's tests revealed that equal variances were assumed for all measures (see Appendix 5, Table 5.2.30).

Differences between geography classrooms at time 2

School 1. ANOVA results by geography classroom at time 2, controlling for prior achievement, are presented in Appendix 5, Table 5.2.8. Similarly to maths, the results show that for the majority of measures, there were no significant differences between geography classrooms following multiple testing correction of $p \leq .000$ ($p = .05/114$). Geography self-perceived ability, geography classroom environment and geography homework feedback did not quite reach significance ($p = .004$, $p = .003$, and $p = .003$ respectively). Levene's tests revealed that equal variances were assumed for most measures apart from homework feedback (see Appendix 5, Table 5.2.31). A significant effect of classroom was found for the following four measures:

Geography Year 5 school achievement time 2. Modest significant differences between classrooms were found for Year 5 school achievement, $F(7,158) = 4.022$, $p < .001$, $\eta_p^2 = .151$, with the highest mean score revealed for class C1e and the lowest for C5se. This was the only significant difference between classes. Following multiple testing correction of $p \leq .000$ ($p = .05/114$), pairwise comparisons revealed students on average, did not differ in their geography achievement at Year 5. Levene's test showed equal variances were assumed for these analyses ($p = .218$).

Geography performance time 2. Modest significant differences between classrooms were found for geography performance, $F(7,156) = 5.209$, $p < .001$, $\eta_p^2 = .189$, with the highest mean score revealed for class C7se and the lowest for C5se. Following multiple testing correction of $p \leq .000$ ($p = .05/114$), pairwise comparisons revealed that students in class C7se on average performed significantly better than students in C5se ($p < .001$), this was the only significant difference between classes. Levene's test showed equal variances were assumed for these analyses ($p = .817$).

***Geography classroom student-teacher relations time 2*.** Modest significant differences between classrooms were found for, $F(7,156) = 3.989$, $p < .001$, $\eta_p^2 = .152$, with the highest mean score revealed for class C3e and the lowest for C5se. Following multiple testing correction of $p \leq .000$ ($p = .05/114$) pairwise comparisons showed that students on average, did not evaluate their student-teacher relations differently as a result of being in a specific class. Levene's test showed equal variances were assumed for these analyses ($p = .082$).

School 2. ANOVA results for school 2 by geography classroom are presented in Appendix 5, Table 5.2.9 and show no significant differences between geography classrooms for all of the measures following multiple testing correction of $p \leq .000$ ($p = .05/114$). Levene's tests revealed that equal variances were assumed for all measures (see Appendix 5, Table 5.2.32).

Maths and geography teacher group differences at time 2

Further analyses were conducted to establish whether patterns of significance found at time 1 persisted at time 2 for teacher groups, when controlling for prior achievement.

Maths and geography teachers. ANOVA results, controlling for prior achievement, are presented in Appendix 5, Table 5.2.10, for maths teachers and Table 5.2.11 for geography teachers. The results show for the majority of measures, no significant differences between maths or geography teacher groupings following multiple testing correction of $p \leq .001$ ($p = .05/57$). Several measures were just below the significance threshold; maths school achievement (p =.005), maths performance (p =.006), number line (p =.002), student-teacher relations (p =.004), and maths environment (p =.004). Levene's tests revealed that equal variances were assumed for most measures apart from maths classroom chaos and geography self-perceived ability (see Appendix 5, Tables 5.2.33 and 5.2.34). The following three measures were significant:

***Maths classroom environment time 2*.** Modest significant differences between teacher groupings were found for classroom environment, $F(5,190) = 4.441$, $p = .001$, $\eta_p^2 = .105$, with the highest mean score revealed for TM4 and the lowest for TM1. Following multiple testing correction of $p \leq .001$ ($p = .05/57$) pairwise comparisons showed that students on average did not evaluate their classroom environment differently if taught by a specific teacher. Levene's test showed equal variances were assumed for these analyses ($p = .140$).

***Maths classroom peer competition time 2*.** Modest significant differences between teacher groupings were found for classroom peer competition, $F(5,188) = 4.270$, $p = .001$, $\eta_p^2 = .102$, with the highest mean score revealed for TM4 and the lowest for TM1. Following multiple testing correction of $p \leq .001$ ($p = .05/57$), pairwise comparisons showed that students did not on average, rate peer competition differently as a result of being taught by a specific teacher. Levene's test showed equal variances were assumed for these analyses (p = .771).

Geography environment time 2. Modest significant differences between teacher groupings were found for geography environment, $F(4,181)$ = 9.918, p < .001, η_p^2= .180, with the highest mean score revealed for TG3 and the lowest for TG2. Following multiple testing correction of $p \leq .001$ (p = .05/57), pairwise comparisons revealed that students studying geography with teacher TG3, on average, rated their geography environment significantly higher in terms of using equipment such as compasses etc., than students studying geography with teachers TG2, TG4 and TG5 (p < .001). Levene's test showed equal variances were assumed for these analyses (p = .755).

Class and teacher group ranking by mean score at time 2

To investigate whether the ranking patterns found at time 1 remained at time 2 after controlling for prior achievement, classes and teacher groups were ranked by their mean scores, from highest to lowest, for all measures that reached significance.

Maths classroom time 2. The results for school 1, controlling for prior achievement, in Appendix 5, Table 5.2.12 show some variability of rank for most classes across the measures. The most consistent class is C8se, which ranked in eighth place for four out of the six measures. Class C6se, which ranked highly previously, ranked in first place for three measures but third, fourth and seventh for the remainder. Some consistency of rank was shown between number line and maths performance, but less was seen between number line and maths achievement. Classroom chaos did not appear to be consistent with maths achievement and performance for most classes apart from C4se and C8se, which ranked low for both (a low score indicates high chaos). Class

C7se, with the least chaos, also ranks low for maths achievement and performance, in sixth and fifth place respectively. For school 2, two measures that fell just below the significance threshold were ranked for parity (see Appendix 5, Table 5.2.13). The classes showed consistency across these measures with C10ce in first and C11ce in third place.

Geography classroom time 2. Similarly to maths class rankings, the rankings for geography controlling for prior achievement, show much variation for the majority of classes across the measures (see Appendix 5, Table 5.2.14). This is apart from C5se, which is consistently in eighth place across all measures. Class C4se and C8se also sit towards the lower ranks and classes C3e and C6se sit in the higher ranks.

Maths teacher time 2. For maths teacher groups, when controlling for prior achievement, there were several measures that fell just below the significance threshold but were included to make comparisons with classrooms and other waves. The results in Appendix 5, Table 5.2.15 show slightly less variation in rank across the measures than for the class groups, with some correspondence but mainly for specific groups. Groups TM4 and TM5 rank towards the top across all the measures and TM2 ranks towards the bottom. TM1 and TM6 sit mainly towards the bottom and TM3 is less consistent. The most consistency was seen across classroom environment, student-teacher relations and peer competition, however we would expect this as student-teacher relations and peer competition are subscales of classroom environment.

Geography teacher time 2. For geography teacher groups, controlling for prior achievement, only geography environment was significant. Appendix 5, Table 5.2.16 shows the rank comparison with school achievement. Complete correspondence is only seen for TG2, ranked in fifth place for both measures, although TG1 and TG5 are ranked similarly across the two measures.

Differences between maths classrooms at time 3

School 1. ANOVA results by maths classroom at time 3, controlling for prior achievement, are presented in Appendix 5, Table 5.2.17. The results show no significant differences between maths classrooms following multiple testing correction of $p \leq .000$ ($p = .05/114$). Four measures that fell just below the significance threshold were maths performance ($p = .002$), classroom environment ($p = .008$), student-teacher relations ($p = .001$), and classroom chaos ($p = .002$)**.** Levene's tests revealed that equal variances were assumed for most measures apart from number line, homework behaviour and maths anxiety (see Appendix 5, Table 5.2.29).

School 2. ANOVA results by maths classroom at time 3, controlling for prior achievement, are presented in Appendix 5, Table 5.2.18. The results show, for the majority of measures differences between maths classrooms following multiple testing correction of $p \leq .000$ ($p = .05/114$). Four measures that fell just below significance were student-teacher relations ($p = .003$), peer competition ($p = .005$), and maths classroom chaos ($p = .001$). Levene's tests revealed that equal variances were assumed for most measures apart from classroom chaos (see Appendix 5, Table 5.2.30). Only one measure reached significance:

Maths classroom environment time 3**.** Moderate significant differences between classrooms were found, $F(2,28) = 13.399$, $p < .001$, $\eta_p^2 = .489$, with the highest mean score revealed for class C10ce and the lowest for C11ce. Following multiple testing correction of $p \leq .000$ ($p = .05/114$), pairwise comparisons revealed that students on average, did not evaluate their environment differently as a result of being in a specific class group. Levene's test showed equal variances were assumed for these analyses (p = .466).

Differences between geography classrooms at time 3

School 1. ANOVA results by geography classroom at time 3, controlling for prior achievement, are presented in Appendix 5, Table 5.2.19. The results show, for the majority of measures, no significant differences between geography classrooms following multiple testing correction of $p \leq .000$ (p = .05/114). Four measures that fell just below significance were geography performance (p = .008), student-teacher relations (p = .002), geography classroom chaos (p = .007), and geography anxiety (p = .004). Levene's tests revealed that equal variances were assumed for most measures apart from classroom environment and classroom chaos (see Appendix 5, Table 5.2.31).The following two measures were significant:

***Geography classroom environment time 3*.** Modest significant differences between classrooms were found for classroom environment, $F(7,144) = 4.282$, $p < .001$, $\eta_p^2 = .172$, with the highest mean score revealed for class C6se and the lowest for C8se. Following multiple testing correction of $p \leq .000$ (p = .05/114), this was the only pairwise comparison that reached significance with students in class C6se rating their classroom environment significantly better than students in class C8se ($p < .001$). However, Levene's test showed unequal variances for these analyses (p = .005) with the smallest variance revealed for class C6se (0.29) and the largest for class C4se (1.42).

***Geography environment time 3*.** Moderate significant differences between classrooms were found for geography environment, $F(7,139) = 7.051$,

$p < .001$, $\eta_p^2 = .262$, with the highest mean score revealed for class C6se and the lowest for C8se. Following multiple testing correction of $p \leq .000$ (p = .05/114), pairwise comparisons revealed that students in class C6se on average, rated their geography environment in terms of using equipment such as compasses etc., significantly higher than students in classes C4se, C7se and C8se ($p < .001$). Levene's test showed equal variances were assumed for these analyses ($p = .511$).

School 2. ANOVA results by geography classroom at time 3, controlling for prior achievement, are presented in Appendix 5, Table 5.2.20. The results show, for all measures, no significant differences between geography classrooms following multiple testing correction of $p \leq .000$ (p = .05/114). Geography classroom chaos fell just below the significance threshold (p = .001). Levene's tests revealed that equal variances were assumed for all measures (see Appendix 5, Table 5.2.32).

Maths and geography teacher group differences at time 3

ANOVA results, controlling for prior achievement, are presented in Appendix 5, Table 5.2.21 for maths teachers and Table 5.2.22 for geography teachers. The results show for the majority of measures, no significant differences between maths or geography teacher groupings following multiple testing correction of $p \leq .001$ (p = .05/57). Measures falling just below the significance threshold were: maths classroom environment ($p = .005$), geography anxiety ($p = .003$). Levene's tests revealed that equal variances were assumed for most measures except number line, maths classroom environment, maths anxiety, geography classroom environment, geography peer competition, and geography anxiety (see Appendix 5, Tables 5.2.33 and

5.2.34). The following five measures were significant:

Maths performance time 3. Modest significant differences between teacher groupings were found for maths performance, $F(5,176) = 4.271$, $p = .001$, $\eta_p^2 = .108$, with the highest mean score revealed for TM5 and the lowest for TM6. Following multiple testing correction of $p \leq .001$ ($p = .05/57$), pairwise comparisons revealed that students on average did not perform differently as a result of being taught by a specific maths teacher. Levene's test showed equal variances were assumed for these analyses ($p = .141$).

Maths student-teacher relations time 3**.** Modest significant differences between teacher groupings were found for student-teacher relations, $F(5,177) = 4.748$, $p = .001$, $\eta_p^2 = .118$, with the highest mean score shown for TM5 and the lowest for TM1. Following multiple testing correction of $p \leq .001$ ($p = .05/57$), pairwise comparisons revealed that students did not evaluate their student-teacher relations differently as a result of being taught by a specific maths teacher. Levene's test showed equal variances were assumed for these analyses ($p = .174$).

Geography classroom environment time 3**.** Modest significant differences between teacher groupings were found for classroom environment, $F(4,179) = 6.026$, $p < .001$, $\eta_p^2 = .119$, with the highest mean score observed for TG3 and the lowest for TG1. Pairwise comparisons revealed that students studying geography with teacher TG3 on average, rated their classroom environment significantly higher than students studying with teachers TG1, TG2 and TG4 ($p < .001$), following multiple testing correction of $p \leq .001$ ($p = .05/57$). Levene's test showed unequal variances for these analyses (p = .008), with the smallest variance shown for teacher TG3 (0.29) who teaches one class and the largest for teacher TG5 (1.30) who teaches two classes.

***Geography student-teacher relations time 3*.** Modest significant differences between teacher groupings were found for student-teacher relations,, $F(4,179) = 5.339$, $p = .001$, $\eta_p^2 = .107$, with the highest mean score shown for TG3 and the lowest for TG1. Following multiple testing correction of $p \leq .001$ ($p = .05/57$), pairwise comparisons revealed that students studying geography with teacher TG3 on average, rated their student-teacher relations significantly higher than students studying with teachers TG1 and TG2 ($p < .001$). Levene's test showed equal variances were assumed for these analyses ($p = .182$).

***Geography environment time 3*.** Moderate significant differences between teacher groupings were found for geography environment, $F(4,173) = 11.925$, $p = .001$, $\eta_p^2 = .216$, with the highest mean score revealed for TG3 and the lowest for TG2. Following multiple testing correction of $p \leq .001$ ($p = .05/57$), pairwise comparisons revealed that students studying geography with teacher TG3 on average, rated their geography environment in terms of using equipment such as compasses etc., significantly higher than students studying with teachers TG2, TG4 and TG5 ($p < .001$). Levene's test showed equal variances were assumed for these analyses ($p = .572$).

Class and teacher group ranking by mean score at time 3

Maths classroom time 3. The stringent multiple testing correction of $p \leq .000$ ($p = .05/114$) meant that for school 1, when controlling for prior achievement, no significant effect of maths classroom was observed for any measure. Four measures that fell just below the significance threshold were ranked for parity with geography and the previous waves. The results in Appendix 5, Table 5.2.23 show some correspondence of rank for specific classrooms across the measures. For example, classes C8se and C5se are ranked in seventh and eighth places, respectively, consistently for all the measures. Class C4se is ranked either fifth or sixth across the measures and C6se is ranked second and third. Class C2e ranks third to first across the four measures.

For school 2, the results in Appendix 5, Table 5.2.24, when controlling for prior achievement, show just one measure reached significance following multiple testing correction of $p \leq .000$ ($p = .05/114$). Four measures that were just below the threshold were included for comparison. Almost complete correspondence for all classes and all measures is observed with C10ce and C9ce in first and second place, respectively, for all measures apart from maths homework behaviour when their positions switch. Class C11ce remains in third place throughout.

Geography classroom time 3. The results for school 1, when controlling for prior achievement, in Appendix 5, Table 5.2.25 show just two measures reached significance following multiple testing correction of $p \leq .000$ ($p = .05/114$), therefore, four additional measures were included for comparison. Less consistency is revealed compared with maths classroom although classes C4se, C5se and C8se rank towards the lower end across the measures. Higher levels of geography anxiety are revealed for C4se and C8se as they rank first and third place respectively on this measure (high score indicates high anxiety). Class C6se sits towards the higher ranks for all measures apart from anxiety, and is consistent with previous waves.

For school 2, when controlling for prior achievement, no measures reached significance following multiple testing correction of $p \leq .000$ ($p = .05/114$). For parity and to make comparisons, geography performance and geography classroom chaos were ranked and the results are presented in Appendix 5, Table 5.2.26. Similarly to maths C10ce is first place, C9ce in second and C11ce is in third place across both measures.

Maths teachers time 3. Appendix 5, Table 5.2.27 shows the results for maths teacher group rankings for measures showing a significant effect of teacher, when controlling for prior achievement. Maths classroom environment was included for comparison as it fell just below significance. Some correspondence is observed across the groups with almost complete consistency between classroom environment and student-teacher relations. TM1, TM2 and TM6 sit towards the lower ranks while TM3, TM4 and TM5 are towards the higher ranks, predominantly.

Geography teachers time 3. The results in Appendix 5, Table 5.2.28 show slightly more consistency across teacher groups and measures than for maths teachers when controlling for prior achievement. For example, complete correspondence is observed between classroom environment and student-teacher relations. Four teacher groups showed consistency for three out of four measures. Geography anxiety, included as it fell just below significance, revealed that low anxiety (indicated by a low score) did not necessarily equate to a better classroom environment or student-teacher relations.

5.2 Discussion

The aim of part 5.2 was to explore whether patterns of significant effects and rankings persisted when taking account of prior achievement. If primary school effects are strong, then it would be expected to see a large reduction in effects once primary school achievement was controlled for. When considering the comparison analyses conducted on maths performance at time 1, the expectation appears to be partly fulfilled. The reduction in effect size for classrooms and a slightly stronger effect of current maths teacher group, suggest that primary school was indeed having some influence (see Tables 5.2.1 to 5.2.3. When primary school achievement was controlled for at time 2 however, the classroom differences were largely unchanged apart from significant effects revealed for maths and geography year 5 achievement which did not show significant differences without controlling for prior achievement. It may be that teachers base the students' year 5 grade on their grades from primary school. Slightly more moderate effect sizes were revealed for classrooms, ranging from 15.1% to 22.7% (see Appendix 5, Tables 5.2.6 to 5.2.9). For teacher groups, significant effects were fewer and effect sizes remained modest ranging from 10.2% to 18% (see Appendix 5, Tables 5.2.10 and 5.2.11). These findings are in the opposite direction to those found for maths performance at time 1 when controlling for prior achievement. This suggests that perhaps at time 1, the influence of primary school classroom/teacher is slightly stronger but weakens by time 2.

At time 3 when prior achievement was controlled for, the differentiation from analyses without controls was more pronounced as no significant differences were found between maths classrooms (see Appendix 5, Table 5.2.17 and 5.2.18). Differences were found between geography classrooms for the same two measures that reached significance without controls, but slightly increased effect sizes were shown (17.2% for classroom environment and 26.2% for geography environment) (see Appendix 5, Tables 5.2.19 and 5.2.20). For teacher groups, slightly fewer measures reached significance when controlling for primary school achievement but the effect sizes remained similar.

These findings suggest a very weak influence from primary school extended to time 3 influencing differences between maths classrooms, which disappeared when prior achievement was controlled for. For geography classrooms, the picture is slightly different as with a slight increase in effect sizes, it appears that any influence from primary school was attenuating differences between geography classrooms, which emerge marginally more strongly when prior achievement is removed.

In comparing the ranking patterns across analyses with and without controlling for primary school achievement, Table 5.2.4 and 5.2.5 show for maths performance at time 1 little or no change for most classes, but two changed by three or four places. Similarly for teachers, only two classes changed. Teacher TM5 who was in first place without controlling for prior achievement, teaches just one class, C6se who is regularly seen in high ranks. The change of position for TM1 is interesting as this teacher was previously in the lower ranks. This teacher also teaches one class, C9ce in school 2.

At time 2 when controlling for prior achievement the ranking patterns for maths classrooms are largely unchanged from analyses without controls, apart from C8se who is ranked in eighth place more frequently (Appendix 5, Table 5.2.12 and 5.2.13). Maths environment is the same with and without controls. Maths achievement saw a few differences, but it seems to be for specific classes. Some change occurred for classes C9ce and C10ce as they switched places across the analyses. For geography, the resemblance between the analyses, is similar to that shown for maths classroom apart from class C8se who, consistent with rankings for maths measures, ranks slightly lower when prior achievement is controlled for (Appendix 5, Table 5.2.14). Geography achievement, like maths achievement, shows some differences in ranking for some classes. For the teacher groups, very little change is observed across maths and geography for most groups (Appendix 5, Table 5.2.15 and 5.2.16). Only one or two groups rank differently TM4 who teaches one class and TM6 who teaches four classes including C8se. Remarkably, the maths classroom environment measures show agreement across the two analyses, suggesting that with or without controlling for primary school teacher, the relationships within the maths classrooms are the same.

When controlling for primary school achievement at time 3 there is some change in rank compared to without controls, but the changes are slight, only one or two ranks apart. Slightly more variation is seen across analyses for maths performance and geography performance, with marginally more change for geography classrooms (see Appendix 5, Tables 5.2.23 to 5.2.26). For the teacher groups, the ranks remain relatively unchanged apart from one or two groups (see Appendix 5, Tables 5.2.27 and 5.2.28). Maths classroom environment is the same across analyses, but there is one change for student-teacher relations. Geography teacher groups remain largely unchanged, which is striking considering there appears to be slightly more variation for classrooms.

In comparing the ranking positions between controlling for primary school achievement at time 2 and time 3 and ranking positions at time 1 (Chapter 4, pp. 139, 140, 141, 154 and 155) the findings show considerably less agreement across the classroom rankings. Especially when comparing across primary school achievement and year 5 achievement. For geography teacher groups the ranks also appear dissimilar apart from teacher TG3 who ranks highly across assessments and analyses. Maths teacher groups show agreement for classroom environment measures but are dissimilar for maths performance.

Overall, fewer significant differences between classrooms and teacher groups when controlling for prior achievement suggest some influence from primary school achievement. The similarity in rank positions across analyses with and without controlling for prior achievement for the majority of classes at time 1, and 2 also advocates some impact of primary school achievement for most classes and teacher groups at time 2. The absence of effects for maths and slightly less agreement between rankings which emerged at time 3 suggests a weakening of any primary school influence by time 3 for maths classrooms and teacher groups. There may however, be some impact for geography classrooms at time 3 as the slight strengthening of effects when prior achievement was controlled. There may also be some impact for just a few specific classes. When controlling for prior achievement, the rankings overall are dissimilar from ranking positions at time 1, which might imply support for the idea of primary school influence. However, the analyses without controls also showed little or no agreement with rankings at time 1. It may be that as the academic year progresses, the classes/teacher groups, overall, loosen their ties with the primary school classroom. Because some classes and teacher groups appear to change in response to the different analyses, it suggests that some influence may remain for them. Primary school achievement, however, does not fully account for potential differences between children, as grading is crude (3, 4, or 5), and many children decrease grades as material becomes harder in secondary school. It also could mean that the observed effects after controlling for primary school achievement still carry effects of the ‘class’ in terms of ethos, relationships etc. Taken together, the findings suggest a weakening influence from primary school, yet, its origins are undefined.

5.3. Classroom Differences in the UK Sample

Classroom differences at time 1

ANOVA were repeated by classroom in the UK sample, without controlling for prior achievement, to enable a comparison with an education system that employs formal selection processes. Students in the UK sample were selected into their maths classes on the basis of prior ability. At this assessment wave there were six classes distributed across 3 levels of ability (2 classes at each level) numbered 1 to 3 with 1 being the highest ability classes. Large achievement differences would be expected between high ability and low ability classes. However, predictions are less clear regarding classroom environment variables such as student-teacher relations, or classroom chaos. Means, standard deviations (SD) and N for all assessed variables by classroom, are presented in Appendix 5, Tables 5.3.1 to 5.3.4 for time 1; Tables 5.3.7 and 5.3.8 for time 2; and Tables 5.3.10 and 5.3.11 for time 3.

Maths classroom time 1. ANOVA results by maths classroom at time 1, are presented in Appendix 5, Table 5.3.1 and 5.3.2. The results show that for most measures, there was no significant differences between maths classrooms following multiple testing correction of $p \leq .001$ ($p = .05/98$). Maths homework feedback fell just below the significance threshold ($p = .003$). Levene’s tests revealed that equal variances were assumed for most measures except self-perceived ability, homework feedback and homework total scale (see Appendix

5, Table 5.3.13). A significant effect of classroom was found for the following six measures:

Maths school achievement time 1. Maths school achievement is assessed every 6 weeks, the timing of this test coincided with the time 1 data collection. As expected, large significant differences between classrooms were found for maths school achievement, $F(5,150) = 42.479$, $p < .001$, $\eta_p^2 = .586$, with the highest mean score revealed for class C1 and the lowest for C3. Following multiple testing correction of $p \leq .001$ ($p = .05/98$), pairwise comparisons revealed that students in class C1, on average, had significantly higher achievement than students in classes C2, C3, R2 and R3 ($p < .001$); students in class R1, on average had significantly higher achievement than students in classes C2, C3, R2 and R3 ($p < .001$); and students in class C3 on average had significantly lower achievement than students in classes C2 and R2 ($p < .001$). Levene's test showed equal variances were assumed for these analyses ($p = .825$).

Maths performance time 1. Also as expected, moderate significant differences between classrooms were found for maths performance, $F(5,152) = 17.565$, $p < .001$, $\eta_p^2 = .366$, with the highest mean score revealed for class C1 and the lowest for C3. Following multiple testing correction of $p \leq .001$ ($p = .05/98$), pairwise comparisons revealed that students in class C1 on average, had significantly higher maths performance than students in classes C3, R2 and R3 ($p < .001$); students in class R1 on average, had significantly higher maths performance than students in classes C3, R2 ($p < .001$), and R3 ($p = .001$); and students in class C3 on average, had significantly lower maths performance than students in class C2 ($p = .001$). Levene's test showed equal variances were assumed for these analyses ($p = .079$).

Number line time 1. Modest significant differences between classrooms were found for the number line task, $F(5,150) = 4.426$, $p < .001$, $\eta_p^2 = .12.9$, with the lowest (optimum) mean score revealed for class R1 and the highest for R3. Following multiple testing correction of $p \leq .001$ ($p = .\ 05/98$), pairwise comparisons showed that students on average did not differ in their number estimation as a result of being in a specific class. Levene's test showed equal variances were assumed for these analyses ($p = .377$).

Maths anxiety time 1. Modest significant differences between classrooms were found for maths anxiety, $F(5,151) = 4.201$, $p = .001$, $\eta_p^2 = .122$, with the highest mean score (indicating high anxiety) revealed for class C3 and the lowest for R1. Indicating, as might be expected, the highest levels of maths anxiety for students in the lowest ability class and the least for students in the highest ability class. Following multiple testing correction of $p \leq .001$ ($p = .\ 05/98$), pairwise comparisons showed that on average, students' levels of maths anxiety did not differ as a result of being in a specific class. Levene's test showed equal variances were assumed for these analyses ($p = .188$).

Cognitive ability test time 1. As expected, large differences between maths classrooms were found for the cognitive ability test scores, $F(5,120) = 27.816$, $p < .001$, $\eta_p^2 = .537$, with the highest mean score revealed for class R1 and the lowest for C3. Following multiple testing correction of $p \leq .001$ ($p = .\ 05/98$, pairwise comparisons revealed as might be expected, that students in a high ability class, R1, on average, had significantly higher cognitive ability scores than students in classes C2, C3 and R2 ($p < .001$); students in a high ability class, C1, on average, had significantly higher cognitive ability scores than students in classes C2, C3 and R2 ($p < .001$); and students in a lowest ability class C3 on average, had significantly lower cognitive ability scores than students in class R2 (p = .001). Levene's test showed equal variances were assumed for these analyses (p = .142).

Theories of intelligence time 1. Modest differences between maths classrooms were found for theories of intelligence, $F(5,147) = 4.359$, $p = .001$, $\eta_p^2 = .129$, with the highest (optimum) mean score revealed for class C1 and the lowest for R3. Following multiple testing correction of $p \leq .001$ (p = . 05/98), pairwise comparisons showed that this was the only comparison that reached significance indicating that students in class C1, on average were more likely to be incremental theorists than students in class R3 ($p < .001$). this suggests that students in the high ability class (C1) were more likely to view intelligence as something that is changeable rather than fixed, compared to students in the lower ability class R3. However, Levene's test showed unequal variances for these analyses (p = .027), with the smallest variance shown for class C1 (0.40) and the largest shown for class C2 (1.12).

Perceptions of academic and socioeconomic status: school grades time 1. As expected, modest differences between maths classrooms were found for perceptions of socio-economic status regarding school grades, $F(5,138) = 4.177$, $p = .001$, $\eta_p^2 = .131$, with the highest mean score revealed for class R1 and the lowest for R3. Following multiple testing correction of $p \leq .001$ (p = . 05/98), pairwise comparisons revealed that students in high ability class R1 on average, rated their position in the school regarding their school grades significantly higher than students in lower ability class C3 ($p < .001$). However, Levene's test showed unequal variances for these analyses (p = .022) with the smallest variance shown for class R1 (0.28) and the largest for class C1 (1.04)**.**

Geography classroom time 1. ANOVA results by geography classroom at time 1, are presented in Appendix 5, Tables 5.3.3 and 5.3.4. The results show that for most measures, there was no significant differences between geography classrooms following multiple testing correction of $p \leq .001$ (p = .05/98). Geography classroom environment fell just below the significance threshold ($p = .006$). Levene's tests revealed that equal variances were assumed for most measures except geography performance (see Appendix 5, Table 5.3.14). A significant effect of classroom was found for the following two measures:

Geography student-teacher relations time 1. Modest significant differences between classrooms were found for student-teacher relations, $F(5,122) = 4.835$, $p < .001$, $\eta_p^2 = .165$, with the highest mean score revealed for class 7C and the lowest for 7R. Following multiple testing correction of $p \leq .001$ (p = . 05/98), pairwise comparisons showed that this was the only comparison that reached significance ($p < .001$), indicating that students in class 7C on average, rated their student-teacher relations significantly higher than students in class 7R, Levene's test showed equal variances were assumed for these analyses ($p = .495$).

Geography classroom chaos time 1. Modest significant differences between classrooms were found for, $F(5,121) = 4.920$, $p < .001$, $\eta_p^2 = .169$, with the highest mean score (high score indicates low chaos) revealed for class 7C and the lowest (high chaos) for 7A. Following multiple testing correction of $p \leq .001$ (p = . 05/98), pairwise comparisons showed that this was the only comparison that reached significance ($p = .001$), indicating that students in class 7C on average, rated their classrooms' chaos level significantly lower than students in class 7A. Levene's test showed equal variances were assumed for these analyses ($p = .589$).

Class ranking by mean score at time 1 for UK sample

The classes were ranked by their mean scores, form highest to lowest, to assess any consistency of class ranking across the significant measures. Appendix 5, Tables 5.3.5 and 5.3.6 show the rankings separately for maths and geography classrooms. Measures that fell just below significance were also included for comparison. Cognitive ability test was also included with geography classroom rankings for comparison even though no significant effect was shown.

Maths classroom time 1. The results for maths classroom in Appendix 5, Table 5.3.5 show correspondence across the measures between class ranking and their streaming. This is evident in the complete correspondence shown across maths school achievement and maths performance for all classes. The higher ability classes sit in/towards the top ranks and lower ability classes sit in/towards the lower ranks. This is also reflected in the rankings for cognitive ability test which is used to stream maths classes when students start secondary school. It appears from the rankings that the lower ability classes rate their homework feedback more highly than the higher ability classes. The lower ability classes also show higher levels of maths anxiety (a high score indicates high anxiety) compared to the higher ability classes. Perceptions of theories of intelligence and academic status (school grades) are observed to be lower for the lower ability classes.

Geography classroom time 1. The results for geography classroom in Appendix 5, Table 5.3.6 show substantial consistency of rank across the significant measures considering these classes are not streamed for ability. We might expect consistency across classroom environment and student-teacher relations as student-teacher relations is a subscale of classroom environment.

However, the rankings for classroom chaos are highly consistent with the other two measures. The cognitive ability test was compared to see whether there was any correspondence between ability and streaming processes with these measures. No consistency was shown at all between the test and the significant measures suggesting no influence of ability on the geography classroom environment rankings.

Classroom differences at time 2 for the UK sample

Maths classroom time 2. ANOVA results by maths classroom at time 2 presented in Appendix 5, Tables 5.3.7 show for most measures, no significant differences between maths classrooms following multiple testing correction of $p \leq .001$ ($p = .05/98$). Maths homework total score fell just below the significance threshold ($p = .008$). Levene's tests revealed that equal variances were assumed for most measures except school achievement, maths performance, self-perceived ability, homework behaviour, homework total scale and maths environment (see Appendix 5, Table 5.3.13). A significant effect of classroom was found for the following five measures:

Maths school achievement time 2. Maths school achievement is assessed every 6 weeks, the timing of this test coincided with the time 2 data collection. As expected, very large significant differences between classrooms were found for maths achievement, $F(5,147) = 113.791$, $p < .001$, $\eta_p^2 = .795$, with the highest mean score revealed for class C1 and the lowest for C3. Following multiple testing correction of $p \leq .001$ ($p = .05/98$), pairwise comparisons revealed that students in class C1 had on average, significantly higher maths achievement than students in all other classes including R1 which was the second highest ($p < .001$); students in class R1 had significantly higher maths achievement than students in all the other classes apart from C1(p < .001); students in class C3 on average had significantly lower maths achievement than all classes apart from R3 ($p < .001$). However, Levene's test showed unequal variances for these analyses ($p = .001$) with the least variance shown for class C2 (0.12) and the most for class C1 (0.38).

Maths performance time 2. As might be expected, moderate significant differences between classrooms were found for maths performance, $F(5,149) = 17.827$, $p < .001$, $\eta_p^2 = .374$, with the highest mean score revealed for class C1 and the lowest for C3. Following multiple testing correction of $p \leq .001$ (p = . 05/98), pairwise comparisons revealed that students in class C1 on average, had significantly higher maths performance than students in classes C3, R2 and R3 ($p < .001$); students in class R1 also had significantly higher maths performance than students in classes C3, R2 and R3 ($p < .001$). However, Levene's test showed unequal variances for these analyses ($p = .033$) with the least variance shown for class R1 (0.30) and the most for class C2 (1.12)

Number line time 2. Moderate significant differences between classrooms were found for the number line task, $F(5,146) = 6.678$, $p < .001$, $\eta_p^2 = .186$, with the lowest (optimum) mean score revealed for class C1 and the highest for R3. Following multiple testing correction of $p \leq .001$ (p = . 05/98), pairwise comparisons showed that students in class R3 performed significantly worse at number estimation than students in classes C1 ($p < .001$) and R1 (p = .001). Levene's test showed equal variances were assumed for these analyses ($p = .464$).

Maths homework feedback time 2. Contrary to expectation, modest significant differences between classrooms were found for homework feedback, $F(5,146) = 4.271$, $p = .001$, $\eta_p^2 = .128$, with the highest mean score revealed for class R3 and the lowest for R1. This indicates that students in the lower ability class (R3) rated their homework feedback the most favourably. Following multiple testing correction of $p \leq .001$ (p = . 05/98), pairwise comparisons showed that on average students did not rate their homework feedback differently as a result of being in a specific class. Levene's test showed equal variances were assumed for these analyses ($p = .113$).

Maths anxiety time 2. Modest significant differences between classrooms were found for maths anxiety, $F(5,145) = 4.466$, $p = .001$, $\eta_p^2 = .133$, with the highest mean score revealed for class C3 and the lowest for C1. Following multiple testing correction of $p \leq .001$ (p = . 05/98), pairwise comparisons showed that students on average did not differ in levels of anxiety as a result of being in a specific class. Levene's test showed equal variances were assumed for these analyses ($p = .364$).

Geography classroom time 2. ANOVA results by geography classroom at time 2 are presented in Appendix 5, Table 5.3.8. No significant differences between geography classrooms were observed for any of the measures following multiple testing correction of $p \leq .001$ (p = .05/98). Levene's tests revealed that equal variances were assumed for all measures (see Appendix 5, Table 5.3.14).

Class ranking by mean score at time 2 for the UK sample

Maths classroom time 2. The measures demonstrating a significant effect of maths classroom were ranked by their mean scores, highest to lowest, and are shown in Appendix 5, Table 5.3.9. Measures just below the significance threshold were also included for comparison. The results show almost complete consistency across the measures and between class groups. The rankings fall completely in line with ability streaming. Similarly to time 1, lower ability classes (e.g. R3) appear to rate their homework feedback more highly than the high ability classes. Compared to time 1, the rankings here show higher consistency.

Geography classroom time 2. No measures were ranked for geography classrooms as no significant effects were shown.

Classroom differences at time 3 in the UK sample

Maths classroom *time 3.* ANOVA results by maths classroom at time 3 presented in Appendix 5, Table 5.3.10 show for most measures, no significant differences between classrooms following multiple testing correction of $p \leq .001$ ($p = .05/98$). Several measures that fell just below the significance threshold were student-teacher relations ($p = .006$), peer competition ($p = .005$), maths environment ($p = .004$), and maths anxiety ($p = .003$). Levene's tests revealed that equal variances were assumed for most measures except maths anxiety (see Appendix 5, Table 5.3.15). A significant effect of classroom was found for the following four measures:

Maths performance time 3. As expected, large significant differences between classrooms were found, $F(7,156) = 21.851$, $p < .001$, $\eta_p^2 = .495$, with the highest mean score revealed for class C1 and the lowest for C4. Following multiple testing correction of $p \leq .001$ ($p = .\ 05/98$), pairwise comparisons revealed that students in class C1 on average, had significantly better maths performance than students in classes C3, C4, R2, R3 and R4 ($p < .001$); students in class R1 on average, also had significantly better maths performance than students in classes C4, and R2 ($p = .001$), and classes C3, R3 and R4 ($p < .001$); students in class C2 also had significantly better maths performance than students in class R4 ($p = .001$). Levene's test showed equal variances were assumed for these analyses (p = .724).

Number line time 3. Expected moderate differences between classrooms were found for the number line task, $F(7,154) = 7.012$, $p < .001$, $\eta_p^2 = .242$, with the lowest (optimum) mean score revealed for class C1 and the highest for R4. Following multiple testing correction of $p \leq .001$ (p = . 05/98), pairwise comparisons revealed that students in class R4 on average performed significantly worse at number estimation than students in classes C1 and R1 (p = .001). Levene's test showed equal variances were assumed for these analyses (p = .605).

Maths classroom environment time 3*.* Contrary to expectation**,** modest significant differences between classrooms were found for classroom environment, $F(7,150) = 4.193$, $p < .001$, $\eta_p^2 = .164$, with the highest mean score revealed for class C1 and the lowest for R3. No pairwise comparisons reached significance following multiple testing correction of $p \leq .001$ (p = . 05/98).

Maths homework feedback time 3. Against expectation, modest significant differences between classrooms were found for homework feedback, $F(7,153) = 3.830$, $p = .001$, $\eta_p^2 = .149$, with the highest mean score revealed for class C3 and the lowest for R1. Following multiple testing correction of $p \leq .001$ (p = . 05/98), pairwise comparisons revealed this as the only significant difference ($p < .001$), indicating that students in class C3 on average, rated their homework feedback more favourably than students in class R1. Levene's test showed equal variances were assumed for these analyses (p = .201).

Geography classroom time 3. ANOVA results by geography classroom at time 3 are presented in Appendix 5, Table 5.3.11. No significant differences between geography classrooms were observed for any of the measures following multiple testing correction of $p \leq .001$ ($p = .05/98$). Levene's tests revealed that equal variances were assumed for most measures except homework total scale (see Appendix 5, Table 5.3.16).

Class ranking by mean score at time 3 for the UK sample

Maths classroom. The measures demonstrating a significant effect of maths classroom were ranked by their mean scores, highest to lowest, and are shown in Appendix 5, Table 5.3.12. Measures just below the significance threshold were also included for comparison. The results show less consistency across the measures and class groups than in earlier waves. There is some correspondence between maths performance and the number line task. There is less consistency than expected between classroom environment, student-teacher relations and peer competition, considering the latter two are subscales of classroom environment. There is some correspondence between rank and streaming as higher ability classes sit towards the higher ranks and the lower ability classes sit towards the lower ranks for most measures.

Geography classroom. No measures were ranked for geography classrooms as no significant effects were shown.

5.3 Discussion

The aim of part 5.3 was to investigate the research question of whether potential significant effects and ranking patterns found in the Russian sample are similar to any potential effects found in the UK sample. Particularly as the UK sample are subject to selection processes for their maths classrooms. The comparison of the results between the UK and Russia present an important contribution to understanding the nature of the effects. If most of the effects are a product of students' ability and other characteristics, then it should be expected that differences between classes would show large effects and high ability classes would repeatedly rank highly.

Indeed, substantial effects were found for maths classroom measures at time 1 ranging from 12.2% for maths anxiety, to 58.6% for maths school achievement (see Appendix 5, Table 5.3.1 and 5.3.2). This effect size for school achievement is completely consistent with the allocation of students to classrooms based on their ability and reflects ability selection rather than other classroom influences. Also in line with streaming processes, stronger effects were shown for measures of achievement and performance rather than classroom environment measures. The number of measures that revealed a significant effect of class was only slightly higher than in the Russian analyses and in many cases, reflected those at time 2 for Russia. Additional measures that reached significance in the UK were maths anxiety, theories of intelligence and perceptions of school grades. The cognitive ability test completed by the UK students also showed a large effect of classroom (53.7%). This test was used to select students into appropriate classrooms based on their ability. Primary school achievement, however, did not show a significant effect of classroom which is interesting when many schools use these results for streaming purposes.

At time 2 differences between maths classrooms showed larger effects, ranging from 12.8% for maths anxiety to 79.5% for maths achievement (Appendix 5, Table 5.3.7). The same measures reached significance apart from theories of intelligence and perceptions of school grades which were not assessed at time 2. Similarly to Russia, there were fewer significant effects at time 3 although effect sizes matched those revealed in earlier waves (14.9% to 45.5%) (Appendix 5, Table 5.3.10).

According to previous research, strong correlations have been shown across different subjects, for example between maths, English and science due to intelligence and other educationally relevant traits pertinent across subjects (e.g. Krapohl et al., 2014). For example, if selection for maths classrooms is on IQ then effects might be expected only for maths classrooms and not for geography classrooms where no selection processes are employed. However, given the correlations between domains significant effects of classroom might be expected for geography classrooms in performance regardless of their mixed ability and unselected status. In fact, the findings showed significant differences between geography classrooms for just a few measures. An effect of classroom was found at time 1 only for student-teacher relations and classroom chaos with effect sizes of 16.5% and 16.9% respectively (Appendix 5, Table 5.3.3).

Similarly to the Russian sample, unequal variances were observed between classrooms for several measures that demonstrated a significant effect of classroom. As with the Russian classrooms, no consistent pattern was observed for increased variance for specific classrooms. This finding indicates a degree of caution should be applied in interpreting these particular results.

Overall, the number of significant effects is similar across the two samples and consistent for some of the measures. The differences observed between the UK sample's maths classrooms, as expected, did exhibit much stronger effects than those seen in the Russian sample. Of interest is the significant effect of maths classroom on maths anxiety only in the UK sample at time 1, and 2, and just below significance at time 3. We might also expect along with this a significant effect of self-efficacy. Although this was not observed, there is a significant effect of maths classroom on perception of school grades, where students place themselves on a ladder in relation to the whole school. The lower ability class groups are at the lower ranks on this measure. The effect on subject anxiety does not appear for geography in the UK sample. In the Russian sample geography anxiety was significant at time 3 only for teacher groups without controlling for prior achievement; maths anxiety was not close. It may be that ability selection is influencing maths anxiety in the UK.

The ranking patterns in the UK sample show more consistency for maths classrooms in rankings than the Russian maths and geography classrooms, as would be expected given the streaming processes (Appendix 5, Tables 5.3.5, 5.3.9 and 5.3.12). Predictably, high ability classes, C1 and R1, occupy the higher ranks and lower ability classes, C3/C4 and R3/R4, occupy the lower ranks. There is still some variation, especially at time 1 but much less so than observed for the Russian sample in parts 5.1 and 5.2. The rankings for UK sample geography classrooms at time 1 show some correspondence across measures (Appendix 5, Table 5.3.6). However, as most of the measures relate to classroom environment, it is feasible that the peer group or teacher may be in influence here. The lack of correspondence with cognitive ability test rankings clearly signifies no influence of ability or selection processes for geography.

The findings for the Russian sample lie somewhere between the UK maths and geography classrooms. The Russian classrooms should be more comparable with the geography classes in the UK sample being that they are also mixed ability. However, the significant effects and ranking consistency across classrooms in the Russian sample, are far greater than those shown for UK geography classes. The ranking patterns are not as clearly defined, however, as those seen for the UK maths classrooms. There appears to be some correspondence with the UK geography classroom, given the significant effects for classroom environment variables in the UK at time 1. However, it is striking that the Russian geography classroom measures continue to show a significant effect throughout the academic year, whereas in the UK, any effect diminishes by time 2. The continuing effect for Russia may be due to variation in student ability and/or implicit selection processes. It may also be due to prior achievement. Equally, the influence may be due to a stronger effect of teacher/classroom, extending from having the same peer group and primary school teacher for so many years. Furthermore, there may indeed be influences from peers and/or influences from current subject teachers.

5.4 Associations Between Teacher Characteristics, Classroom Environment and Academic Outcomes

Further analyses were conducted to disentangle any influence of primary school teacher on secondary school classroom environment and outcomes in the Russian sample. Bivariate correlations between primary school teacher characteristics and measures demonstrating a significant effect of maths and geography classroom without controlling for prior achievement were conducted for time 1, time 2 and time 3 (see Appendix 5, Tables 5.4.2 to 5.4.7). These analyses were also repeated with maths and geography teacher characteristics to establish potential associations between current class teachers' characteristics and maths and geography outcomes (see Appendix 5, Tables 5.4.8 to 5.4.13). Means and standard deviations (SD) of teacher characteristics are presented in Table 5.4.1.

In reporting these results, a causal influence is implied in certain places. With longitudinal associations, it can be assumed that the relationship might be causal. However, given the correlational nature of these analyses, it is recognised that other factors may also contribute or may fully explain the associations.

The associations between teacher characteristics, for example, between teacher self-efficacy in student engagement and teacher emotional ability. shown for primary school, maths, and geography teachers will not be discussed. The teacher characteristics are only used as a grouping variable to assign categories to students as the teacher sample alone (*N*=17) is not adequate to make inferences regarding any significant associations between teacher characteristics.

Associations between primary school teacher characteristics and maths and geography classroom measures

The results show a weak influence of primary school teacher/classroom on future maths and geography classroom measures across the first year of secondary education. The influence of primary school teacher/classroom on maths classroom measures reduced considerably as the academic year progressed, although some influence remained for geography classroom at time 3.

Maths classroom.

Time 1. Primary school maths achievement was also included at time 1 to explore potential associations (see Appendix 5, Table 5.4.2). A negative association revealed that when primary school maths achievement was lower, primary school teacher emotional ability was higher, albeit weakly (r = -.216, $p \leq .01$). This relationship was concurrent, but a prospective positive association was also revealed between primary school teacher emotional ability and maths performance at time 1 (r = .172, $p \leq .05$). Primary school achievement also associated with student teacher relations at time 1 (r = .155, $p \leq .05$). The results also indicate a less chaotic maths classroom environment at time 1 as a consequence of primary school teacher characteristics. Weak positive associations were revealed between maths classroom chaos (a high score indicates low chaos) at time 1 and the following primary school teacher characteristics: years of experience (r = .204, $p \leq .01$); self-efficacy in student engagement (r = .156, $p \leq .05$); and self-efficacy in instructional strategies (r = .274, $p \leq .01$). These results imply some influence of primary school achievement and teacher on some measures for maths classrooms at time 1.

Time 2. The results presented in Table 5.4.3 also maintain some influence of primary school teacher that extends to maths classroom at time 2. Although no significant associations were revealed between primary school teacher characteristics and maths achievement at time 2, a weak positive association was found between primary school teacher experience and maths performance at time 2 (r = .170, $p \leq .05$). Weak associations were also revealed between number line and primary school teacher experience (r = -.264, $p \leq .01$), emotional ability (r = -.157, $p \leq .05$), and self-efficacy in instructional abilities (r = -.284, $p \leq .01$). A low score is optimum for the number line task so the negative relationships actually indicate higher ability levels associated with better primary school teacher characteristics. Similarly to time 1, primary school teacher influence extended to maths classroom chaos at time 2 with a weak positive association revealed between maths classroom chaos and primary school teacher self-efficacy in instructional strategies ($r = .220$, $p \leq .01$). However, no association is revealed between classroom chaos and teacher self-efficacy in student engagement, instead, a weak negative relationship is shown with teacher self-efficacy in classroom management ($r = -.147$, $p \leq .05$). This suggests that higher primary school teachers' self-efficacy in classroom management was associated with a more chaotic maths classroom environment at time 2 (low score indicates high chaos). The weak negative association revealed between primary school teacher emotional ability and homework feedback at time 2 ($r = -.263$, $p \leq .01$) suggests that having a primary school teacher with higher emotional ability, may lead a student to have, comparatively, less encouraging perceptions of future teacher feedback. Together these results suggest a weak influence of primary school teacher on some measures of maths classroom environment at time 2. For other measures, such as maths achievement, associations would likely be stronger with current subject teacher.

Time 3. The results in Appendix 5, Table 5.4.4 show that any influence of primary school teacher on maths classroom has reduced substantially at time 3. The only significant association is between maths classroom chaos and primary school teacher self-efficacy in instructional strategies, albeit weakly ($r = .252$, $p \leq .01$). This suggests, however, a continuing influence of primary school teacher on maths classroom atmosphere at least up to time 3, the beginning of the next academic year.

Geography classroom.

Time 1. The results in Appendix 5, Table 5.4.5 show associations between primary school teacher characteristics and geography classroom measures at time 1. Geography primary school achievement, also included here to explore potential associations, correlated weakly and positively with primary school teacher self-efficacy in student engagement (r = .153, $p \leq .05$), and primary school teacher experience (r = .207, $p \leq .01$). These associations were contemporaneous; however, prospective relationships were also shown that suggest the influence of primary school teacher on later geography classrooms. Weak positive associations were revealed between primary school teacher experience and geography classroom environment at time 1 (r = .207, $p \leq .01$), student-teacher relations at time 1(r = .207, $p \leq .01$), and geography classroom chaos at time 1(r = .207, $p \leq .01$). Weak positive associations were also shown between self-efficacy in instructional strategies and geography classroom environment at time 1 (r = .203, $p \leq .01$), student-teacher relations at time 1 (r = .183, $p \leq .01$), and geography classroom chaos at time 1 (r = .207, $p \leq .01$).

Time 2. The results for geography classroom at time 2 in Appendix 5, Table 5.4.6 show a negative relationship between geography performance at time 2 and primary school teacher emotional ability (r = -.233, $p \leq .01$), self-efficacy in student engagement (r = -.238, $p \leq .01$), and self-efficacy in classroom management (r = -.235, $p \leq .01$). These results suggest that students whose primary school teacher had higher emotional ability and confidence had lower geography performance at time 2. Weak positive associations were revealed between geography environment at time 2 and primary school teacher experience (r = .253, $p \leq .01$), and self-efficacy in instructional strategies (r =

.215, $p \leq .01$). These results imply a weak influence of primary school teacher on geography classroom measures at time 2.

Time 3. The results at time 3 in Appendix 5, Table 5.4.7 indicate a continuing influence of primary school teacher on geography classroom into the beginning of the next academic year. Primary school teacher self-efficacy in classroom management weakly and negatively associated with geography performance at time 3 ($r = -.185$, $p \leq .05$), and student-teacher relations at time 3 ($r = -.155$, $p \leq .05$). These results show that students whose primary school teacher had higher confidence in classroom management had lower geography performance and worse student teacher relations at time 3. This may be a consequence of an authoritarian style of primary school classroom management impacting negatively on later performance and classroom environment. Conversely, primary school teacher instructional strategies associated positively, although weakly, with student-teacher relations at time 3 ($r = .165$, $p \leq .05$), and geography environment at time 3 ($r = .290$, $p \leq .01$). Geography environment at time 3 also associated positively with primary school teacher experience ($r = .253$, $p \leq .01$), primary school teacher emotional ability ($r = .179$, $p \leq .05$), and primary school teacher self-efficacy in student engagement ($r = .178$, $p \leq .05$). These results show some influence from primary school teacher on geography classroom outcomes at time 3. The negative associations observed are unexpected, but may give some insight into the complex nature of classroom influence.

Associations between current maths teacher characteristics and maths classroom measures

The results for maths teacher characteristics also showed weak associations at time 1, time 2 and time 3. Similarly to primary school teacher characteristics, the number of associations between maths teacher characteristics and maths classroom measures reduced at time 3.

Time 1. Appendix 5, Table 5.4.8 shows the results between maths teacher characteristics and maths classroom measures at time 1, which also includes primary school maths achievement. Weak negative associations were revealed between primary school maths achievement and current maths teacher emotional ability (r = -.181, $p \leq .05$), maths teacher self-efficacy in student engagement (r = -.141, $p \leq .05$), and maths teacher self-efficacy in classroom management (r = -.193, $p \leq .01$). These results indicate that lower primary school achievement (prior) may have led to higher perceived emotional ability and self-efficacy in the current teacher, although the process underlying this association is unclear. The results also show concurrent negative associations between maths teacher characteristics and maths classroom measures. Weak negative associations were shown between maths performance at time 1 and maths teacher self-efficacy factors: student engagement (r = -.243, $p \leq .01$), and instructional strategies (r = -.236, $p \leq .01$). Weak negative associations were also observed between student-teacher relations at time 1 and maths teacher emotional ability (r = -.298, $p \leq .01$). Maths classroom chaos at time 1 also negatively associates with self-efficacy factors: student engagement (r = -.142, $p \leq .05$); and instructional strategies (r = -.166, $p \leq .05$). These results suggest that maths teacher perceived emotional ability and self-efficacy is higher when students’ maths performance is lower and when students perceive student-teacher relations and classroom chaos as more negative (a high score indicates low chaos). One potential explanation is that teachers with higher self-perceived abilities possess some other characteristics that are perceived more negatively by the students or lead to lower outcomes. However, this hypothesis is not supported by previous research that found positive associations between teacher self-efficacy and student characteristics (Tschannen-Moran, & Hoy, 2007). Another explanation may be that students with low performance (i.e. lower ability students) have been assigned stronger, more confident teachers.

Time 2. At time 2 year 5 school achievement was also included as the effect of classroom differed when controlling for prior achievement. The results in Appendix 5, Table 5.4.9 show negative relationships between current maths teacher characteristics and maths classroom measures at time 2. Interestingly, no associations were revealed, positive or negative, between maths achievement at time 2 and maths teacher characteristics. Maths performance at time 2 weakly and negatively associated with maths teacher emotional ability ($r = -.146$, $p \le .05$), and maths teacher self-efficacy factors: student engagement ($r = -.223$, $p \le .01$); instructional strategies ($r = -.191$, $p \le .01$); and classroom management ($r = -.144$, $p \le .05$). Associations were also revealed between number line at time 2 and maths teacher self-efficacy factors: student engagement ($r = .317$, $p \le .01$); instructional strategies ($r = .276$, $p \le .01$); and classroom management ($r = .243$, $p \le .01$). Bearing in mind that a low score is optimum for the number line task, these positive results suggest that students whose maths teachers had more confidence had lower performance on the task. Similarly, students with lower maths performance at time 2 had maths teachers with high self-perceived emotional ability and confidence. Maths classroom environment at time 2 also negatively associated with maths teacher emotional ability (*r* = -.200, *p* ≤ .01), suggesting that students of maths teachers' with high self-perceived emotional ability rated their classroom environment unfavourably. Maths classroom environment at time 2 also positively associates with maths teacher self-efficacy factors: student engagement (*r* = .180, *p* ≤ .05); and classroom management (*r* = .184, *p* ≤ .01). Weak negative associations were also revealed between maths classroom chaos at time 2 and maths teacher experience (*r* = -.159, *p* ≤ .05), and self-efficacy in instructional strategies (*r* = -.154, *p* ≤ .01) suggesting higher levels of classroom chaos in relation to more years of experience and higher levels of confidence in instructional strategies. Although these maths teacher characteristics negatively associated with the study measures, they did not associate with school maths achievement. This may be because the school achievement measure is crude (only 3 categories used, corresponding to satisfactory, good and excellent). It may be that differences in teacher characteristics do not associate with school maths achievement. The achievement score is derived by the teacher in question, consequently the teacher may perceive more control in this outcome, and therefore not be subject to performance pressure.

Time 3. The results for maths teacher characteristics and maths classroom measures at time 3 in Appendix 5, Table 5.4.10 show fewer significant associations than for time 1 and time 2. Weak positive associations were revealed between maths teacher self-efficacy in classroom management and maths classroom environment at time 3 (*r* = .192, *p* ≤ .01), and student-teacher relations at time 3 (*r* = .218, *p* ≤ .01). Weak negative associations were shown between maths classroom environment at time 3 with maths teacher experience (*r* = -.149, *p* ≤ .01), and maths teacher emotional ability (*r* = -.156, *p*

≤ .01). Similarly, student-teacher relations at time 3 also negatively associated with maths teacher experience (*r* = -.147, *p* ≤ .05), and maths teacher emotional ability (*r* = -.153, *p* ≤ .05). These results suggest that in classes whose teachers had more experience and higher self-perceived emotional ability, students evaluated their classroom environment and student-teacher relations more negatively. One potential explanation for this phenomenon may be teacher burn out, whereby following years of teaching, emotional exhaustion may lead a teacher to treat their students indifferently (Maslach, Jackson, & Leiter, 1996).

Associations between current geography teacher characteristics and geography classroom measures

The results for geography teacher characteristics and geography classroom measures show a higher number of significant associations across the three assessments, compared to maths teacher characteristics and maths classrooms.

Time 1. The results for time 1 in Appendix 5, Table 5.4.11 show a weak negative association between geography teacher experience and geography classroom chaos at time 1 (*r* = -.153, *p* ≤ .05). As a high chaos score indicates low chaos, this result suggests that students rated their classrooms as more chaotic if their teacher had more experience. However, students who rated their classrooms as less chaotic had teachers with higher self-perceived teacher self-efficacy factors: student engagement (*r* = .298, *p* ≤ .01); instructional strategies (*r* = .267, *p* ≤ .01); and classroom management (*r* = .320, *p* ≤ .01). Geography primary school achievement was also estimated here. Unlike maths, however, geography at primary level may be significantly different to secondary school geography. Primary school achievement was shown to negatively associate with current geography teacher emotional ability (r = -.271, $p \leq .01$), suggesting that students with lower prior achievement had teachers with higher levels of self-perceived emotional ability. Students with higher geography primary school achievement had current geography teachers with higher self-perceived teacher self-efficacy factors: student engagement (r = .221, $p \leq .01$); instructional strategies (r = .279, $p \leq .01$); and classroom management (r = .271, $p \leq .01$). Similar results were revealed for geography performance at time 1 and geography teacher characteristics: a negative association with geography teacher emotional ability (r = -.259, $p \leq .01$), and positive associations with self-efficacy factors: student engagement (r = .172, $p \leq .05$); instructional strategies (r = .169, $p \leq .05$); and classroom management (r = .196, $p \leq .05$). This pattern of results was also repeated for geography student-teacher relations at time 1: negatively associated with geography teacher emotional ability (r = -.318, $p \leq .01$), and positively associated with self-efficacy factors: student engagement (r = .198, $p \leq .05$); instructional strategies (r = .220, $p \leq .01$); and classroom management (r = .233, $p \leq .01$). This pattern of results also reflects the negative association between geography teacher emotional ability and self-efficacy factors. Together, they indicate that teachers with higher self-perceived self-efficacy had students with higher performance and better student-teacher relations at time 1; and students with lower performance and worse student-teacher relations had teachers with higher self-perceptions of emotional ability. Students with higher primary school geography achievement also had current geography teachers with higher self-perceived self-efficacy.

Time 2. Year 5 geography achievement was also estimated at time 2 as different effects were found for this construct with and without controlling for prior achievement. The results in Appendix 5, Table 5.4.12 show a different pattern to the one found for maths. Geography achievement positively associates with teacher self-efficacy factors: student engagement ($r = .256$, $p \leq .01$); instructional strategies ($r = .214$, $p \leq .01$); and classroom management ($r = .216$, $p \leq .01$). Self-efficacy factors also positively associated with geography performance at time 2: student engagement ($r = .267$, $p \leq .01$); instructional strategies ($r = .381$, $p \leq .01$); and classroom management ($r = .291$, $p \leq .01$); also with student-teacher relations at time 2: student engagement ($r = .258$, $p \leq .01$); instructional strategies ($r = .197$, $p \leq .01$); and classroom management ($r = .252$, $p \leq .01$); and more moderately with geography environment at time 2: student engagement ($r = .423$, $p \leq .01$); instructional strategies ($r = .391$, $p \leq .01$); and classroom management ($r = .417$, $p \leq .01$). It may be that teachers of higher achieving students, perceived their own level of self-confidence highly as a consequence of their students doing well. It may be though that with so few teachers, one or two teachers are influencing the pattern of results. The geography environment measure is likely to associate with the teacher self-efficacy factors given the nature of the questions, e.g., ‘How often do you solve geography problems with a partner or in small groups?’ or ‘How often do you work with objects like rulers, compasses, atlases, or maps?’ The apparent growth in the strength of the relationship between time 1 and time 2 may reflect teachers’ increased use of implementing these teaching practices. Of interest is geography teacher experience which shows weak negative associations with student-teacher relations at time 2 ($r = -.164$, $p \leq .05$), and geography environment at time 2 ($r = -.189$, $p \leq .05$), suggesting that students of teachers with more years of experience rated their student-teacher relations and geography environment unfavourably. Negative associations were also revealed between geography teacher emotional ability and geography performance at time 2 (r = -.286, $p \leq .01$), and with geography environment at time 2 (r = -.224, $p \leq .01$).

Time 3. The results for time 3 in Appendix 5, Table 5.4.13 show a similar pattern to time 1 and time 2 with only slightly fewer associations than previously. Geography teacher experience negatively associated with geography environment at time 3 (r = -.267, $p \leq .01$), suggesting teachers with more years of experience made less use of these teaching practices, which seemed to be used more by teachers with fewer years of experience. It may be that teacher burnout, as a result of many years teaching is influencing the lower use of these teaching practices. It may be that that these teaching practices are used more now as a result of changes in teacher training and teachers were not encouraged to use them previously. Similarly to previous waves, geography teacher emotional ability negatively associated with geography performance at time 3 (r = -.258, $p \leq .01$), and this time with geography classroom environment at time 3 (r = -.240, $p \leq .05$). Interestingly, student-teacher relations which would be expected to associate with teacher emotional ability showed no significant association despite being a subscale of classroom environment. These results suggest that students with lower geography performance rated their classroom environment unfavourably and had teachers with high self-perceptions of emotional ability. Positive associations were again revealed between geography teacher self-efficacy factors: and geography performance at time 3: student engagement (r = .222, $p \leq .01$); instructional strategies (r = .305, $p \leq .01$); and classroom management (r = .226, $p \leq .01$); and with student-teacher relations at time 3: student engagement (r = .221, $p \leq .01$); instructional strategies (r = .198, $p \leq .01$); and classroom management (r = .224, $p \leq .01$); and with geography environment at time 3:student engagement (r = .402, $p \leq .01$); instructional strategies (r = .298, $p \le .01$); and classroom management (r = .390, $p \le .01$). Geography classroom environment was included this time as it also showed a significant effect of classroom at time 3, and similarly to student-teacher relations, positive associations were revealed with self-efficacy factors: student engagement (r = .250, $p \le .01$); instructional strategies (r = .272, $p \le .01$); and classroom management (r = .269, $p \le .01$). These results imply fairly stable relationships between geography teacher characteristics and geography classroom environment measures across the academic year.

5.4 Discussion

The aim of part 5.4 was to investigate the research question of whether teacher characteristics associated with measures of classroom environment, performance and achievement. These analyses were conducted in combination with the mediation analyses in part 5.5 in an attempt to disentangle effects from teachers (primary and current subject teachers) which are confounded by class groups. Primary school achievement was also included to separate effects from primary school influences. If primary school teacher influences were strong, then it would be expected that their teacher characteristics would associate with classroom measures and performance. Likewise, if current subject teacher influences were stronger, then stronger associations would be expected.

Overall, the findings signify a weak influence from the primary school classroom that extended across the first year of secondary education. Weak relationships, at around .3 maximum, were revealed between primary school achievement and maths and geography classroom measures. Primary school teacher characteristics also associated with the classroom measures. Some associations, for example, between geography classroom student-teacher relations and primary teacher self-efficacy factors, associated positively and extended to time 3 (Appendix 5, Table 5.4.7). More associations, however, were observed between current subject teachers and the classroom measures. These were weak to moderate, around .4 maximum, suggesting a stronger influence of current subject teacher. However, these associations reduced between maths classroom measures and current maths teacher by time 3. Some associations were negative, for example, between maths performance at time 2 and current maths teacher emotional ability; and with self-efficacy factors (Appendix 5, Table 5.4.9). It is possible this indicates a helpful response from the teacher towards a struggling student. It may also indicate that stronger teachers are assigned to teach groups of weaker students. Compared to maths teacher, geography teacher characteristics associated more frequently with classroom measures and extended further across the academic year. This was also reiterated with the associations between geography teacher self-efficacy factors and geography year 5 achievement at time 2, which were not replicated for year 5 maths achievement and teacher characteristics. It may be that teacher characteristics are less important when it comes to teaching maths as opposed to geography. However, caution should be applied when interpreting these findings as the sample of teachers was small. It may be one or two teachers influencing the results in each domain. Future research will address this in a larger sample.

5.5 Mediating Relationships Between Teacher Characteristics, Classroom Environments And Academic outcomes

The bivariate correlations between teacher characteristics and maths and geography outcomes presented in Appendix 5, Tables 5.4.2 to 5.4.13 also show, for some measures, associations across several variables. Several associations are observed between primary school teachers and/or achievement and secondary school classroom measures. To explore any mediating effects from primary school teacher characteristics/students' primary school achievement and current class teacher characteristics, 19 path analyses were conducted using simple mediation. Multi-mediation models were not conducted due to the extent of multi-collinearity between the teacher characteristic variables (Preacher & Hayes, 2008). Significant mediating effects are reported using bias-corrected bootstrap confidence intervals.

The relationship between maths performance time 1 (T1) and maths classroom chaos (T1) mediated by maths teacher self-efficacy (SE): in student engagement; and instructional strategies

As research suggests that a more orderly classroom leads to better maths performance (Opdenakker, & Damme, 2001), it is expected that this relationship might also be influenced by teacher self-efficacy in student engagement and instructional strategies. Figure 2 shows the reduction in effects between maths classroom chaos at time 1 and maths performance at time 1 in two separate models where they were mediated by maths teacher self-efficacy in student engagement and maths teacher self-efficacy in instructional strategies, respectively. Specific indirect effects were significant with self-efficacy in student engagement as the mediator at 95% CI (0.005 - 0.070). The model was a good fit to the data, AIC = 1878.02; BIC =1915.79; $\chi^2(229)= 1.270$, $p = 0.26$; RMSEA = 0.034; CFI = 0.99; TLI = 0.93; SRMR = 0.021. With maths teacher self-efficacy in instructional strategies as the mediator, specific indirect effects were significant at 99% CI (0.001 - 0.095). This model was also a good fit to the data, AIC = 1761.44; BIC =1799.21; $\chi^2(229)= 1.363$, $p = 0.24$; RMSEA = 0.04; CFI = 0.99; TLI = 0.91; SRMR = 0.02.

The independent relationships between each of the self-efficacy factors with maths classroom chaos and maths performance were negative, suggesting that teachers with high self-perceived teacher self-efficacy had students with lower maths performance and/or who rated their classrooms as more chaotic. However, the direct relationship between chaos and performance was positive (recall, high chaos score indicates low chaos) suggesting better performance in a more orderly classroom. In both models, however, self-efficacy factors positively mediated the relationships, suggesting that teachers with high self-perceived teacher self-efficacy had students who rated their classrooms as unchaotic and had better maths performance. It may be that as teacher confidence in student engagement and instructional strategies improved, the classroom became more orderly, students were more engaged and as a consequence, maths performance improved.

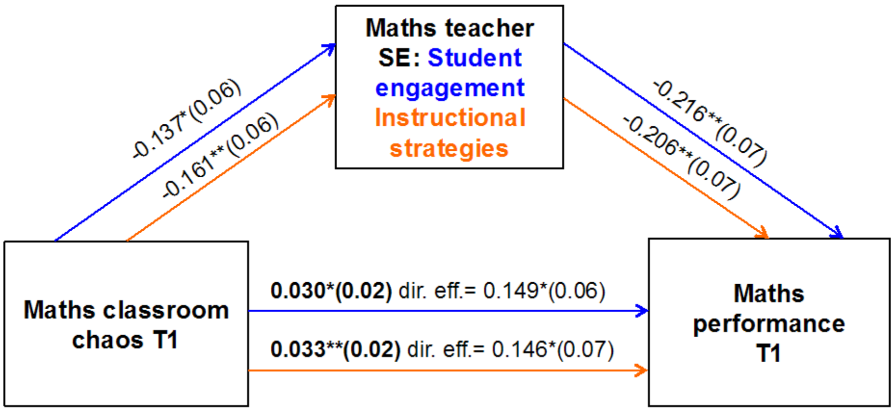

Figure 5.5.1. Summary of two separate, simple mediation models. For each model, maths classroom chaos time (T) 1 was predictor and maths performance time (T) 1 was the dependent variable with maths teacher self-efficacy (SE) in student engagement, and instructional strategies entered separately as the mediators in each model. Paths are colour coded and follow the order for each mediator and standardised beta coefficients are presented with standard errors in parenthesis (* = significant at 95% CI (confidence intervals); **= significant at 99% CI). Paths from the predictor to the dependent variable report beta coefficients for specific indirect effects after mediation (in bold) and direct effects (dir. eff.) before mediation.

The relationship between geography student-teacher relations at time 1 (T1) and geography primary school achievement, mediated by primary school teacher experience

Previous research has shown that achievement associates with better future student-teacher relations (Hughes, Im, & Wehrly, 2014). It is reasonable to expect that the primary school teachers' experience may lead to higher primary school achievement, which in turn may lead to better student-teacher relations. Figure 3 shows the reduction in effects between geography primary school achievement and student-teacher relations at time 1, mediated by primary school teacher experience. Specific indirect effects were significant at 95% CI (0.003, 0.300). The model was a good fit to the data, AIC = 2443.67;

BIC =2481.40; $\chi^2(229)= 0.000$, $p = 0.99$; RMSEA = 0.00; CFI = 1.00; TLI = 1.24; SRMR = 0.00.

As expected, primary school teacher experience positively mediated the relationship between primary school achievement and geography student-teacher relations at time 1, possibly as a result of their students' higher primary school achievement which in turn led to better student-teacher relations.

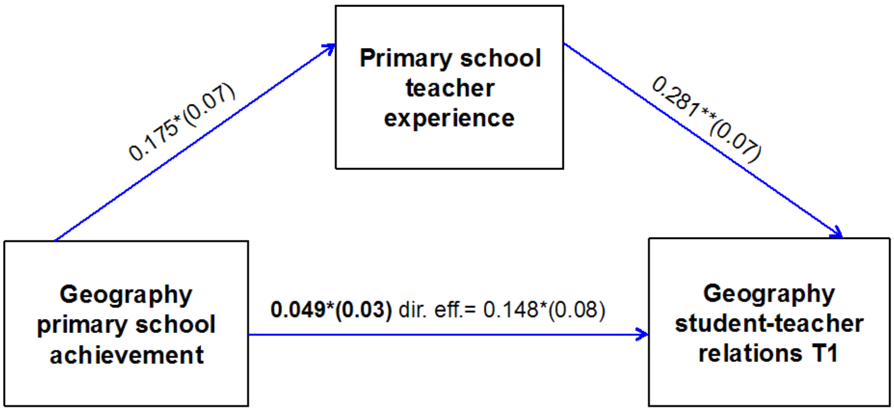

Figure 5.5.2. Summary of the simple mediation model with geography primary school achievement as predictor and student-teacher relations at time (T) 1 as the dependent variable, and primary school teacher experience as mediator. Standardised beta coefficients are presented with standard errors in parenthesis (* = significant at 95% CI (confidence intervals); **= significant at 99% CI). Paths from the predictor to the dependent variable report beta coefficients for specific indirect effects after mediation (in bold) and direct effects (dir. eff.) before mediation.

The relationships between geography classroom chaos time 1 (T1) and geography teacher self-efficacy (SE) factors: student engagement; instructional strategies; and classroom management mediated by geography primary school achievement

It is expected that teachers with higher teacher self-efficacy would have a less chaotic classroom, it might also be expected that student ability, in the form of prior achievement may also influence this relationship (Opdenakker, &

Damme, 2001). Figure 4 shows the reduction in effects between geography classroom chaos at time 1and geography teacher self-efficacy factors: student engagement; instructional strategies; and classroom management, respectively, in three separate models where they were all mediated by geography primary school achievement.

For the model with self-efficacy in student engagement as predictor, specific indirect effects were significant at 95% CI (0.004, 0.099). For the model with self-efficacy in instructional strategies as predictor, specific indirect effects were significant at 99% CI (0.004, 0.151). For the model with self-efficacy in classroom management as predictor, specific indirect effects were also significant at 99% CI (0.001, 0.137). All three models (N = 183) were fully saturated (zero degrees of freedom) and each a perfect fit to the data (Geiser, 2013).

These results suggest that students with high levels of primary school geography achievement had current geography teachers with high self-perceptions of teacher self-efficacy which led to a less chaotic classroom atmosphere.

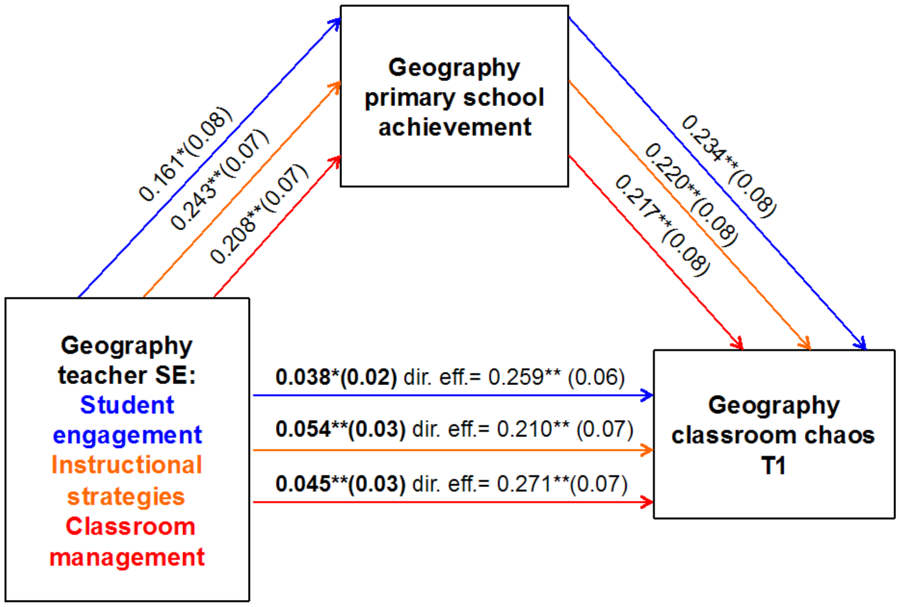

Figure 5.5.3. Summary of three separate, simple mediation models. For each model, the three geography teacher self-efficacy (SE) factors: student engagement; instructional strategies; and classroom management were separate predictors and geography classroom chaos at time (T) 1 was the dependent variable with geography primary school achievement as the mediator in each model. Paths are colour coded and follow the order for each predictor and standardised beta coefficients are presented with standard errors in parenthesis (* = significant at 95% CI (confidence intervals); **= significant at 99% CI). Paths from the predictor to the dependent variable report beta coefficients for specific indirect effects after mediation (in bold) and direct effects (dir. eff.) before mediation.

The relationships between geography year 5 achievement and geography environment time 2 (T2) mediated by geography teacher self-efficacy (SE) factors: student engagement; instructional strategies; and classroom management

The positive associations shown between geography teacher self-efficacy factors, geography environment and geography year 5 achievement suggest the relationship between geography environment and achievement is likely mediated by teacher self-efficacy factors. Figure 6 shows the effects between geography year 5 achievement and geography environment at time 2 in three separate models mediated by geography teacher self-efficacy factors: student engagement; instructional strategies; and classroom management, and geography teacher emotional ability, respectively.

Specific indirect effects were significant with self-efficacy in student engagement as the mediator at 99% CI (0.021, 0.227). This effect increased very slightly from 0.101 to 0.109. The model was a good fit to the data, AIC = 1605.11; BIC =1642.88; $\chi^2(229)= 1.400$, $p = 0.24$; RMSEA = 0.04; CFI = 0.99; TLI = 0.95; SRMR = 0.02.

Specific indirect effects were significant with self-efficacy in instructional strategies as the mediator at 95% CI (0.012, 0.148). The model was a good fit to the data, AIC = 1502.51; BIC =1540.28; $\chi^2(229)= 1.400$, $p = 0.89$; RMSEA = 0.00; CFI = 1.00; TLI = 1.02; SRMR = 0.02.

Specific indirect effects were significant with self-efficacy in classroom management as the mediator at 99% CI (0.003, 0.196). The model was a good fit to the data, AIC = 1651.24; BIC =1689.02; $\chi^2(229)= 0.867$, $p = 0.35$; RMSEA = 0.00; CFI = 1.00; TLI = 1.02; SRMR = 0.02.

The results show that when current geography teachers' self-perceptions of teacher self-efficacy were high, students rated their geography environment highly in use of specific equipment such as rulers, compasses, etc., and in turn, students' year 5 geography achievement was also high.

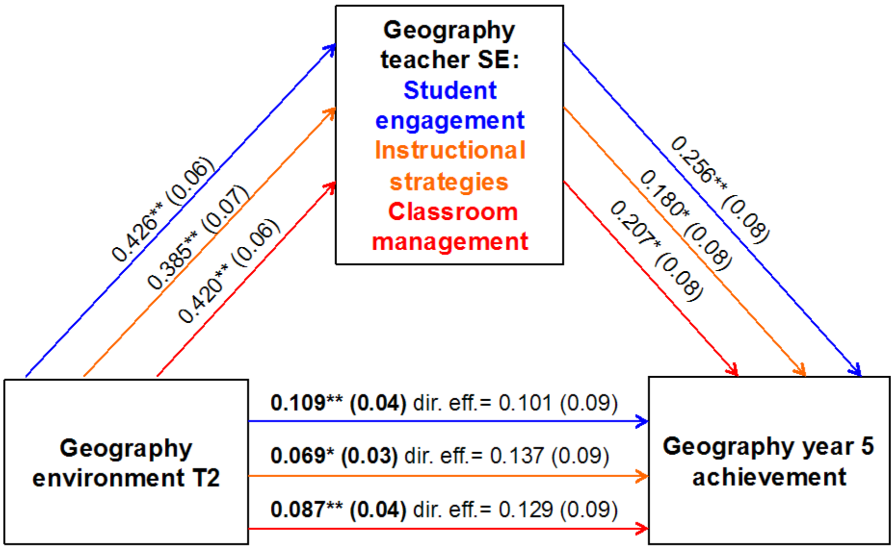

Figure 5.5.4. Summary of three separate, simple mediation models. For each model, Geography environment at time (T) 2 was predictor and geography year 5 achievement was the dependent variable with geography teacher self-efficacy (SE) factors: student engagement; instructional strategies; and classroom management; as the mediators in each model. Paths are colour coded and follow the order for each mediator and standardised beta coefficients are presented with standard errors in parenthesis (* = significant at 95% CI (confidence intervals); **= significant at 99% CI). Paths from the predictor to the dependent variable report beta coefficients for specific indirect effects after mediation (in bold) and direct effects (dir. eff.) before mediation.

The relationships between geography performance time 2 (T2) and geography environment time 2 (T2) mediated by geography teacher self-efficacy (SE) factors: student engagement; instructional strategies; and classroom management, and geography teacher emotional ability

Following the mediating effects of teacher self-efficacy factors between geography environment (time 2) and geography achievement (time 2), it is expected that similar mediating effects of teacher self-efficacy factors will be seen between geography environment and geography performance at time 2. Teacher emotional ability was also expected to mediate the relationship between performance and geography environment at time 2 as it also associated with them. Figure 7 shows the effects between geography performance at time 2 and geography environment at time 2 in four separate models mediated by geography teacher self-efficacy factors: student engagement; instructional strategies; and classroom management, and geography teacher emotional ability, respectively.

Specific indirect effects were reduced and significant with self-efficacy in student engagement as the mediator at 99% CI (0.001, 0.208). The model was a reasonable fit to the data, AIC = 1541.39; BIC =1578.72; $\chi^2(220)= 1.662$, $p = 0.20$; RMSEA = 0.06; CFI = 0.99; TLI = 0.91; SRMR = 0.02.

Specific indirect effects were significant with self-efficacy in instructional strategies as the mediator at 99% CI (0.046, 0.251). This effect increased very slightly from 0.052 to 0.131. The model was a good fit to the data, AIC = 1422.69; BIC =1460.02; $\chi^2(220)= 1.053$, $p = 0.30$; RMSEA = 0.02; CFI = 1.00; TLI = 0.99; SRMR = 0.02.

Specific indirect effects were significant with self-efficacy in classroom management as the mediator at 99% CI (0.020, 0.226). This effect increased very slightly from 0.086 to 0.105. The model was a good fit to the data, AIC = 1581.41; BIC =1618.74; $\chi^2(220)= 1.047$, $p = 0.31$; RMSEA = 0.02; CFI = 1.00; TLI = 0.99; SRMR = 0.02.

Specific indirect effects were reduced and significant with geography teacher emotional ability as the mediator at 99% CI (0.005, 0.138). The model was a reasonable fit to the data, AIC = 950.381; BIC =987.711; $\chi^2(220)= 0.043$, $p = 0.84$; RMSEA = 0.00; CFI = 1.00; TLI = 1.30; SRMR = 0.004.

The results show positive significant mediating effects for each of the four models. Indicating that when current geography teachers' self-perceptions of emotional ability and teacher self-efficacy were high, students at time 2 rated their geography environment high in the use of equipment such as rulers, compasses etc., and working in small groups. This in turn led to better geography performance at time 2.

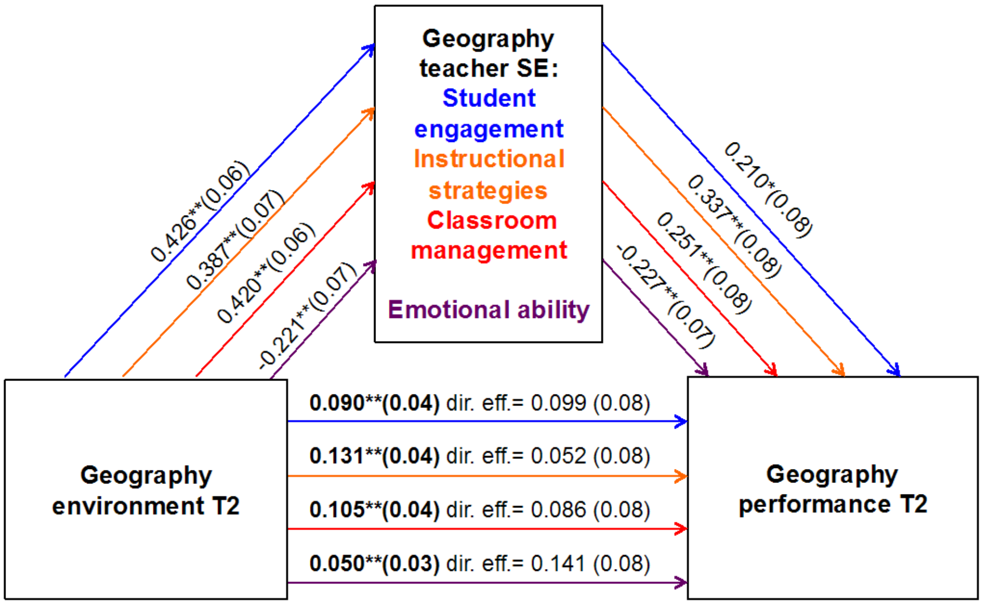

Figure 5.5.5. Summary of four separate, simple mediation models. For each model, Geography environment at time (T) 2 was predictor and geography performance at t time (T) 2 was the dependent variable with geography teacher self-efficacy (SE) factors: student engagement; instructional strategies; and classroom management; and geography teacher emotional ability entered separately as the mediators in each model. Paths are colour coded and follow the order for each mediator and standardised beta coefficients are presented with standard errors in parenthesis (* = significant at 95% CI (confidence intervals); **= significant at 99% CI). Paths from the predictor to the dependent variable report beta coefficients for specific indirect effects after mediation (in bold) and direct effects (dir. eff.) before mediation.

The relationships between geography performance time 3 (T3) and geography environment time 3 (T3) mediated by geography teacher self-efficacy (SE) factors: student engagement; instructional strategies; and classroom management

With the mediating effects of geography teacher self-efficacy factors evident at time 2, it is expected to see similar relationships at time 3. Figure 8 shows the reduction in effects between geography performance at time 3 and geography environment at time 3 in three separate models mediated by geography teacher self-efficacy factors: student engagement; instructional strategies; classroom management, and geography teacher emotional ability, respectively.

Specific indirect effects were significant with self-efficacy in student engagement as the mediator at 95% CI (0.006, 0.150). The model was an acceptable fit to the data, AIC = 1540.97; BIC =1577.99; $\chi^2(214)= 1.858$, $p = 0.17$; RMSEA = 0.06; CFI = 0.98; TLI = 0.87; SRMR = 0.02.

Specific indirect effects were significant with self-efficacy in instructional strategies as the mediator at 99% CI (0.024, 0.190). The model was a reasonable fit to the data, AIC = 1431.20; BIC =1468.23; $\chi^2(214)= 1.540$, $p = 0.21$; RMSEA = 0.05; CFI = 0.98; TLI = 0.91; SRMR = 0.02.

Specific indirect effects were significant with self-efficacy in classroom management as the mediator at 95% CI (0.002, 0.142). The model was a good fit to the data, AIC = 1585.49; BIC =1622.52; $\chi^2(214)= 1.327$, $p = 0.25$; RMSEA = 0.04; CFI = 0.99; TLI = 0.95; SRMR = 0.02.

The results show that when current geography teachers' self-perceptions of teacher self-efficacy were high, students at time 3 rated their geography environments highly in the use of equipment, such as compasses etc., and working in small groups. They also had better geography performance at time 3. It would be expected that teachers confidence in factors such as use of instructional strategies and student engagement would associate with the geography environment measures. It is also reasonable to expect that these associations would also lead to better geography performance.

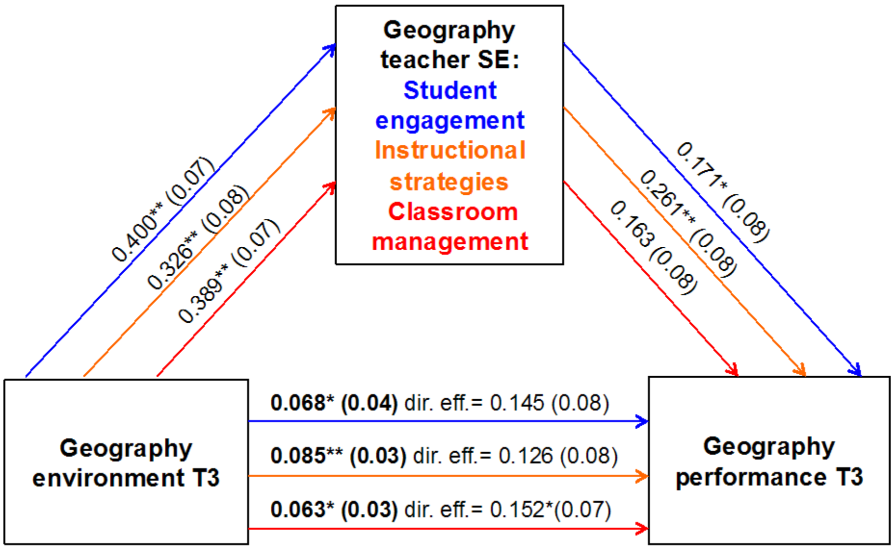

Figure 5.5.6. Summary of three separate, simple mediation models. For each model, Geography environment at time (T) 3 was predictor and geography performance at time (T) 3 was the dependent variable with geography teacher self-efficacy (SE) factors: student engagement; instructional strategies; and classroom management; as the mediators in each model. Paths are colour coded and follow the order for each mediator and standardised beta coefficients are presented with standard errors in parenthesis (* = significant at 95% CI (confidence intervals); **= significant at 99% CI). Paths from the predictor to the dependent variable report beta coefficients for specific indirect effects after mediation (in bold) and direct effects (dir. eff.) before mediation.

The relationships between geography performance time 3 (T3) and geography student-teacher relations at time 3(T3), mediated by geography teacher self-efficacy (SE) factors: student engagement; instructional strategies; and classroom management

Given the mediating relationship of primary school teacher self-efficacy in classroom management with geography performance and student-teacher relations at time 3, and the associations shown for current geography teacher self-efficacy with the geography measures, it is expected that current geography teacher self-efficacy factors would also mediate the relationship between geography performance and student-teacher relations. Figure 9 shows the effects between geography student-teacher relations at time 3 and geography performance at time 3 in three separate models where they were mediated by geography teacher self-efficacy factors: student engagement; instructional strategies; and classroom management respectively. Specific indirect effects were reduced and significant with self-efficacy in student engagement as the mediator at 99% CI (0.003, 0.138). The model was a good fit to the data, AIC = 1542.83; BIC =1579.86; $\chi^2(214)= 1.096$, $p = 0.30$; RMSEA = 0.02; CFI = 1.00; TLI = 0.97; SRMR = 0.02.

Specific indirect effects were significant with self-efficacy in student instructional strategies as the mediator at 99% CI (0.022, 0.170). This effect increased very slightly from 0.065 to 0.080. The model was a good fit to the data, AIC = 1418.70; BIC =1455.73; $\chi^2(214)= 1.091$, $p = 0.30$; RMSEA = 0.02; CFI = 1.00; TLI = 0.98; SRMR = 0.02.

With self-efficacy in classroom management as the mediator, specific indirect effects were reduced and significant at 99% CI (0.001, 0.136). The model was a good fit to the data, AIC = 1584.99; BIC =1622.02; $\chi^2(214)= 0.695$, $p = 0.40$; RMSEA = 0.00; CFI = 1.00; TLI = 1.12; SRMR = 0.02.

The results suggest that better student-teacher relations associate with higher teacher self-efficacy and, in turn, with better geography performance. Because the teacher feels more confident possibly as a consequence of better student-teacher relations, they are able to manage the classroom and engage the students better, perhaps with better instructional strategies, and so geography performance improves as a result.

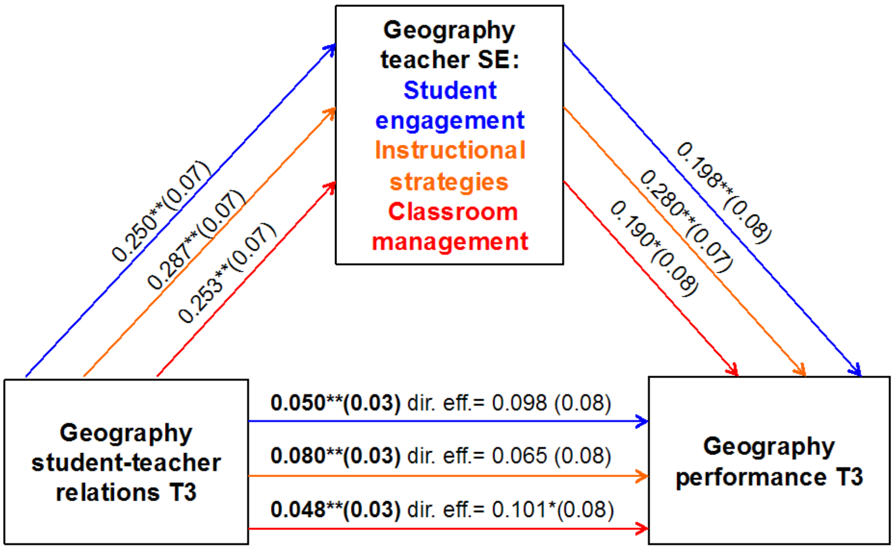

Figure 5.5.7. Summary of three separate, simple mediation models. For each model, Geography student-teacher relations at time (T) 3 was predictor and geography performance at time (T) 3 was the dependent variable with geography teacher self-efficacy (SE) factors: student engagement; instructional strategies; and classroom management entered separately as the mediators in each model. Paths are colour coded and follow the order for each mediator and standardised beta coefficients are presented with standard errors in parenthesis (* = significant at 95% CI (confidence intervals); **= significant at 99% CI). Paths from the predictor to the dependent variable report beta coefficients for specific indirect effects after mediation (in bold) and direct effects (dir. eff.) before mediation.

5.5 Discussion

The aim of part 5.5 was to investigate whether teacher characteristics mediated potential relationships between classroom environment measures and performance/achievement. Primary school achievement was also included to separate effects from primary school influences. If primary school teacher influences were strong, then it would be expected that their teacher characteristics would mediate relationships between classroom measures and performance. Likewise, if current subject teacher influences were stronger, then more mediating effects would be expected.

The findings suggest that relationships between classroom measures and/or performance were indeed mediated by teacher characteristics and separately, by primary school achievement. More relationships were observed for geography classroom measures, however, than for maths. There were also fewer associations in part 5.4 for maths classroom measures which informed these analyses. The models suggest though, that any factor is linked to many other factors and therefore, in itself is unlikely to be causal. For example, self-efficacy factors were mediators in one model but were also predictors in another.

Largely, the results imply that primary school had a weak influence across the first year of secondary education. Firstly, primary school teacher characteristics mediated relationships for geography classroom measures at time 1 (Figure 5.5.2). Secondly, primary school achievement mediated the relationships between geography teacher self-efficacy factors and geography classroom chaos at time 1 (Figure 5.5.3). A greater influence of primary school teacher characteristics was observed for geography classrooms compared to maths classrooms as no such associations or mediating effects were shown for maths classroom measures. It may be that current maths teacher characteristics' had stronger influence as mediating relationships were found for maths classroom measures at time 1 (Figure 5.5.1). Current geography teacher characteristics appeared to have more influence on the classes than maths teacher characteristics as associations and mediating effects were observed at time 1, 2 and 3 (Figures 5.5.4, 5.5.5, 5.5.6, and 5.5.7).

The mediating influence of geography teacher self-efficacy factors on geography environment and geography year 5 achievement (Figure 6) is expected as the measure relates to teaching practices. Equally, it is unsurprising that these relationships were replicated with geography performance at time 2 (Figure 7), and time 3 (Figure 8). Some of the mediating paths slightly increased compared to the direct effects prior to mediation. This may indicate an interesting path, perhaps moderating effects; these increases are very small but they may warrant further exploration in future research with a larger sample.

In the case of primary school achievement as mediator, it is likely that student ability is the mediating factor. However, primary school achievement may be proxy for other contributing factors such as classroom ethos/peer dynamics.

The mediating relationships are positive which indicates a responsive learning environment for students. Where teacher characteristics are mediators, for example in Figures 5.5.4, 5.5.5, 5.5.6, and 5.5.7, this may reflect an evocative process whereby the student characteristic elicits behaviour in the teacher which reinforces teacher self-efficacy. Evocative processes are poorly understood in behavioural research but the behavioural genetic literature investigating family processes has provided numerous examples where genetically influenced characteristics of a child evokes specific behaviours from their parent (e.g. Harlaar, et al., 2008). These studies refer to the relationship between genes in common between parent and child which correlate and increase the environmental influence between them. Although in the classroom, a teacher and student will not have genes in common to correlate and increase the influence, the evocative process is similar in the behavioural sense. For example, if a student appears to learn concepts quickly, it may lead a teacher to increase the pace of instruction in response.

Overall, the findings suggest a weak influence of primary school that extends across the first year of secondary education. They suggest that primary school teacher does indeed play a part, however, so too do current maths and geography teachers. Equally, student prior achievement also has a role. Part 5.5 provides some evidence of specific contributing factors that may confound classroom and teacher group effects.

General Discussion

The main aim of this study was to investigate whether significant differences between classrooms and teachers found at time 1 in Chapter 4 on achievement, performance, classroom environment, student-teacher relations and classroom chaos, in the Russian sample, persisted across time 2 and time 3. In part 5.1, the modest significant differences found at time 2 between classrooms (14.1% to 21.3%) and teacher groups (9.6% to 11.8%) for a few measures, suggest some continuity of effects. However, at time 1, significant differences were present for classroom environment and student teacher relations but they were no longer present for classroom at time 2. Instead, significant differences were shown for number line performance and homework behaviour between classrooms (see Table 5.1.1). Additional differences were also shown between teacher groups for number line performance and peer competition (see Table 5.1.5). Together these findings suggest a slight change from effects found at time 1. Combined with the fewer effects found at time 3, showing modest effects for classrooms (14.2% and 24.6%) and slightly smaller effects for teacher groups (8.5% and 19.6%) (see Tables 5.1.13 to 5.1.19), these findings suggest that any effects found at time 1 are weakening by time 3. However, the unequal variances found between classrooms and teacher groups for some measures potentially interfere with a confident interpretation of results for some analyses.

Ranking patterns were also explored across time 2 and time 3. The greater variation in ranking positions observed for most classes across measures at time 2 and time 3 and less agreement with time 1 suggest more departure from effects at time 1 (see Tables 5.1.7 to 5.1.12 for time 2, and 5.1.20 to 5.1.25 for time 3). With some consistency observed for specific classrooms and teacher groups (e.g. C6se, TM5, TG3 in higher ranks and C5se, TM1 in the lower ranks), the findings indicate that effects may be stronger for specific classes. The overall finding is a weakening effect of classroom/teacher groups observed at time 1 for measures of maths and geography classrooms.

In order to disentangle potential effects from primary school, in part 5.2 the research question was explored whether patterns of significant effects and rankings persisted when taking account of prior achievement. If primary school effects are strong, it would be expected to see a large reduction in effects once primary school achievement was controlled for. The findings of fewer significant differences between classrooms and teacher groups when controlling for prior achievement suggest some influence from primary school achievement.

The agreement in ranking positions for the majority of classes and teacher groups, across analyses with and without controlling for prior achievement at time 2 and 3 might also advocate some impact of primary school achievement at time 2. However, there was little concordance with ranking patterns at time 1 for both sets of analyses. This may indicate a loosening of ties with primary school for the majority of classrooms and teacher groups.

The absence of effects for maths and slightly less agreement for rankings at time 3 with and without controlling for prior achievement also suggests a weakening of any primary school influence for maths classrooms and teacher groups by time 3. There may however, be some impact for geography classrooms at time 3 as a slight strengthening of effects was observed when prior achievement was controlled. There may indeed be some impact for just a few specific classes.

It may be that as the academic year progresses, the influence from the primary school classroom largely loosens for the classes/teacher groups. Equally, some classes and teacher groups demonstrated different effects in response to the different analyses, this suggests that some influence of primary school may remain for them. While primary school achievement is closely linked to the curriculum, it cannot fully account for potential differences between children, as with only three grades (3, 4, or 5) the range is limited. Moreover, many students' grades drop as difficulty of material increases in secondary school. It could also mean that the observed effects after controlling for primary school achievement still carry effects of the ‘classroom ethos’, teacher/peer dynamics etc. Taken together, these findings suggest some influence from primary school however, any effect is weakening as the year progresses and its origins are undefined.

The higher or lower ranking positions observed in the Russian sample for specific classrooms and teacher groups indicates the presence of selection processes, albeit informally. One way to understand the nature of these effects was to compare analyses with results of the UK sample where students were subject to selection processes for their maths classrooms, but not for their geography classrooms. In part 5.3, it was expected that differences between UK maths classes would show large effects and high ability classes would rank highly if effects were a product of students’ ability and other characteristics. Indeed, differences between maths classrooms for a few measures were found across the academic year with substantial effects ranging from 12.2% to 79.5% (see Appendix 5, Tables 5.3.1, 5.3.2, 5.3.7 and 5.3.10). Whereas for geography classrooms, a few differences were found at time 1 only with modest effect sizes ranging from 16.5% and 16.9% (see Appendix 5, Tables 5.3.3, 5.3.4, 5.3.8 and 5.3.11). The ranking patterns for maths classrooms were compatible with ability streaming as high ability classrooms populated the high ranks and low ability classes populated the low ranks (see Appendix 5, Tables 5.3.5, 5.3.9 and 5.3.12). Of interest, was the significant differences between maths classrooms observed in the UK for maths anxiety, theories of intelligence and perceptions of school grades, especially given the ranking patterns observed for these that corresponded with high ability students having lower maths anxiety and higher self-perceptions of their school grades. As with the findings for the Russian sample, unequal variances were observed for some analyses, therefore some caution should also be applied when interpreting these results.

The findings suggest that the Russian sample lie somewhere between the UK maths and geography classrooms. Given the mixed ability nature of the UK geography classrooms, it would be expected that any similarity between the two samples would occur here. Significant effects were indeed found for similar measures (e.g. classroom environment rather than performance), but effects for Russian geography classrooms extended across the academic year to time 3 whereas they diminished by time 2 for the UK. Significant effects and ranking consistency between classrooms in the Russian sample were also much greater than those shown for UK geography classes. The effect sizes were smaller, however and ranking patterns less clearly defined than UK maths classrooms.

Overall, these results suggest that continuing effects for the Russian sample may be due to variation in student ability and/or implicit selection processes. While in Chapter 4, the results suggested a negligible connection between implicit selection processes and learning two languages, there may be some impact from parents who unofficially, manage to obtain a classroom place for their child with a popular teacher. Equally, any influence may be due to a stronger effect of teacher/classroom, extending from having the same peer group and primary school teacher for so many years. Furthermore, there may indeed be influences from peers and/or influences from current subject teachers.

It is apparent that effects from teachers, peers, selection processes and/or variation in student ability and primary school are confounded. Although difficult to tease out, one way to attempt this was to explore relationships between teacher characteristics and measures that revealed a significant effect of classroom, without controlling for prior achievement. If primary school teacher influences were strong, then it would be expected that their teacher characteristics would associate with, and potentially mediate relationships between classroom measures and performance. Likewise, if current subject teacher influences are stronger, then stronger associations and mediating effects would be expected.

In part 5.4, the association analyses suggest a weak influence from primary school that extended across the first year of secondary education. The associations between primary school achievement and maths and geography classroom measures of around .3 suggest some influence from prior achievement and/or primary school classroom (see Appendix 5, Table 5.4.2 to 5.4.7 and 5.4.8 and 5.4.11). Weak associations observed between primary school teacher characteristics and geography classroom measures at time 3 indicate the extent of primary school teacher influence across the academic year. However, in line with expectations, associations between current subject teachers and classroom measures of around .4, imply the slightly stronger impact of current teacher.

The mediation analyses in part 5.5 reiterate the weak influence from primary school with mediating effects observed up to time 2 for achievement and up to time 1 for primary school teacher characteristics. The findings for current subject teachers, however, indicate a responsive learning environment for students which is likely due to evocative processes.

Overall, these findings show that influences from primary school do indeed extend across the first year of secondary education, albeit weakly. While primary school achievement may be proxy for student ability, primary school classroom ethos and/or peer dynamics, the findings recognise the contributions of both primary school and current subject teachers. However, as the sample size of teachers was small, some caution should be applied when interpreting the findings. Characteristics of one or two teachers may be in influence here.

Taken together, the findings of this study do suggest a weak influence from primary school years for the Russian sample that extends across the first year of secondary education. Although some evidence is provided for the contribution of primary school teacher and current subject teacher characteristics, the findings do not precisely clarify the existence of specific effects from peers. Regarding selection processes, perhaps similarly to Toronto parents, the parents were engaging in implicit selection when they chose a school for their child with a more challenging language program (Leonard, 2011). Likewise, at the beginning of primary school, some parents may have obtained a classroom place for their child with a popular teacher. Any form of selection may influence the classroom dynamic and differential effects may emerge across the class and teacher groups depending on student ability (e.g. Burgess et al., 2014; Carmen & Zhang, 2012; Ding & Lehrer, 2007; Guyon et al., 2012; Hattie, 2002; Kelly & Carbonara, 2012; Maaz et al., 2008). It may be that the higher levels of maths anxiety and low perception of school grades found for the lower ability classes in the UK sample are a product of selection processes there. Especially as the students were selected for only one or two subjects, which has been shown to have a detrimental effect on lower ability students' mathematical self-concept (Chmielewski et al., 2013). Alternatively, these findings may be the result of a less favourable classroom climate shown to influence student motivation at this stage of development (Maulana et al., 2013). Either way, these effects warrant further exploration.

It may be that certain teachers' characteristics, particularly in the Russian sample, may be influencing differences seen between classrooms and teacher groups. Table 5.1.7 (Appendix 5) shows levels of teacher experience in relation to the class groups. However, there is no indication that differences between classes are the result of any variation in career length. In fact, all teachers, primary and secondary have a number of years experience. There is also no reason to suggest that teachers with the longest careers are having poorer outcomes. The findings suggest that the students experienced a responsive orderly environment conducive to learning (e.g. Opdenakker, & Damme, 2001; Tschannen-Moran, & Hoy, 2007). The students themselves may have elicited responses from their teachers and/or peers and this may have led to the variation in effects seen for classroom environment and student teacher relations (Maulana et al., 2014).

Strengths And Limitations

The study has several limitations. One limitation is the small sample of teachers which suggests caution should be applied when interpreting results from associations and mediation analyses. These were preliminary analyses on a small sample. As it is difficult to find the balance between a comprehensive and in-depth study and adequate sample size, future research is necessary with a larger sample to further explore these relationships. Despite the small number of teachers, one of the strengths of the study is that these data were more comprehensive than from the UK sample. In the UK, several teachers teach a particular class at different times during the academic year. There is also an issue with temporary teachers covering classes for long periods, sometimes for months. Consequently, the teacher data from the UK were not used to make a comparison with the Russian teacher groups. Another strength is that students in Russia attend one school throughout their education and this enabled the collection of data from the students' primary school teachers. Unfortunately, one limitation is that the timing of the study did not allow data collection from the students regarding their primary school classes which would have helped disentangle primary and secondary school classrooms.

Conclusion

In conclusion, this study shows an attenuated contribution from primary school towards variation in student outcomes across classroom and teacher groups. The effect, however, faded across the first year of secondary education and its origins remain largely undefined. While primary school achievement has been implicated, it not only represents student ability it may also embody primary school classroom ethos, teacher characteristics and peer dynamics. As the comparison with the UK results show, some variation may also be a product of student ability and implicit selection processes. The primary school and current subject teacher characteristics identified as mediating influences between the classroom measures and performance may reflect evocative processes which also rely heavily on student characteristics. These findings suggest that student outcomes, rather than being predominantly influenced by teacher effects, are under multiple influences which should be taken into account by policymakers involved in teacher promotion and employment prospects.

Chapter 6

The development of associations between academic anxiety and performance: a longitudinal cross-cultural investigation

Abstract

A number of studies demonstrated reciprocal associations between academic anxiety for specific school subjects and performance in these subjects. The present study explored the development of associations across one academic year between maths anxiety and maths performance, as well as between geography anxiety and geography performance. Analyses reported in Chapter 5 of this thesis showed some differences between the UK and Russian samples in classroom effects on maths anxiety. In particular, differences between classes were observed in the UK (average effect size 13%), but not the Russian sample. This suggests that associations between anxiety and performance may also differ in the two samples, for example moderated by ability streaming in the UK. The present study therefore investigated whether the longitudinal associations between maths anxiety and maths performance differed in the UK and Russian samples. Using multi-group cross-lagged analyses, associations were investigated within and between the two samples. The results showed that associations for maths developed differently in the two samples. In the Russian sample, prior maths performance negatively predicted later maths anxiety. This may be due to students comparing their performance to weaker and stronger peers in the classroom. In the UK sample, prior maths anxiety negatively predicted later maths performance, possibly as a

consequence of individual differences in maths anxiety combined with a rigorous streaming process that moved some students up or down classes every six weeks following a test. For geography constructs no causal/longitudinal associations were observed between anxiety and performance within both samples. This study shows that longitudinal associations between academic anxiety and academic performance manifested differently cross-culturally, and developed differently between academic subjects. Variation found cross-culturally may be a consequence of dissimilarities in education systems.

Introduction

The findings reported in Chapter 3 of the present thesis showed no significant average or variance differences between the UK and Russian samples for almost all the study measures. However, the analyses in Chapter 5 showed somewhat different patterns between the countries in terms of classroom effects for some measures. For example, maths anxiety showed a significant effect of classroom in the UK only. The effect was shown at each assessment wave with average effect sizes of 13%. Given these differences and the mixed literature regarding the association between anxiety and performance, this relationship warrants further investigation within and between the two samples. In particular, there may be some moderation of the anxiety-performance associations as a function of streaming by ability.

Many individuals experience maths anxiety which translates as feelings of apprehension or nervousness when performing mathematical tasks (Ma & Xu, 2004). Distinct from general anxiety (Hembree, 1990), it is shown to be negatively associated with maths achievement/performance with an average association of *r* = -.30 (e.g. Ashcraft & Moore, 2009; Hembree, 1990). It is suggested, that maths anxiety disrupts maths performance more acutely in a testing situation (e.g. Ashcraft & Moore, 2009). However, results are not consistent across studies and effects have been observed at *r* = -.12 (Gliner, 1987) and *r* = -.60 (Saigh & Kouri, 1983). Associations have also been shown between maths anxiety and an increased negative attitude towards maths (Ashcraft & Moore, 2009), which may lead to avoidance of mathematical content (Hembree, 1990). Reduced participation in mathematical pursuits can have an upstream effect on academic achievement, and in turn, career choices, especially in STEM (science, technology, engineering and mathematics) related fields (Eccles, 2012). Maths anxiety has been shown to persist academically (Ma & Xu, 2004) and also to encroach on everyday numerical tasks such as when checking receipts and change (Ashcraft & Moore, 2009).

Maths anxiety has been shown in adults (Maloney, Risko, Ansari, & Fugelsang, 2010), secondary school students (e.g. Devine, Fawcett, Szucs & Dowker, 2012), and in primary school students (e.g. Vukovic, Kieffer, Bailey, & Harari, 2013; Wu, Barth, Amin, Malcarne, & Menon, 2012). However, it remains poorly understood when exactly it emerges. Studies have shown that maths anxiety occurs in children as young as six years old (Krinzinger, Kaufmann, & Willmes, 2009). The findings are inconsistent, however, regarding associations with maths performance. However, at age 7 an association was observed where maths anxiety associated differentially with different aspects of maths (Vukovic et al., 2013).

One mechanism through which anxiety may affect performance is working memory (e.g. Ashcraft, & Krause, 2007;Beilock, 2008). Individuals with higher maths anxiety may experience reduction in working memory to the extent that there are insufficient resources remaining to execute even relatively undemanding maths problems. Even individuals with high levels of working memory can be affected by maths anxiety (e.g. Vukovic, Kieffer, Bailey, Sean, &Harari, 2013). The suggestion is that these individuals are more susceptible because they usually rely on their working memory to solve maths problems and if suffering from maths anxiety, there is far less capacity left to perform the task (Ramirez et al., 2013).

Recent research has demonstrated around 60% of individual differences in maths anxiety are attributable to non-shared environmental factors – influences that contribute to differences among family members. The study showed no influences from shared environmental factors - influences that contribute to similarities among family members, therefore, the remaining 40% of individual differences are attributable to genetic differences (Wang et al., 2014).

One environmental factor shown to relate to maths anxiety levels is parental support and expectation. An indirect influence was shown between parental support/expectation and higher order mathematics by reducing childrens' maths anxiety. Higher levels of expectation and provision of home learning environment led to lower levels of maths anxiety and higher levels of achievement in 7-8 year old children (Vukovic, Roberts, & Green Wright, 2013). As research has shown environmental influence to be non-shared (Wang et al., 2014), the results from this study suggests that parental support and expectation may be individual-specific, whereby the parent responds to the characteristics of the child (Plomin& Bergman, 1991). Within one family, parental response may differ across their offspring, depending on the child’s characteristics (Reiss et al.,1995). Perceptions of the classroom learning environment have also demonstrated an association with maths anxiety (Taylor & Fraser, 2013). Given the anticipation of performance in a maths class (e.g. solving a maths problem in front of class) and the prospect of unfavorable evaluation by peers and teachers, classroom environment is likely to be an important factor in the relationship between maths anxiety and performance (Hopko, McNeil, Zvolensky, & Eifert, 2002). Having a high level of maths anxiety in class, has been shown to not only hinder performance but also to impede learning ability, particularly in children with higher levels of working memory (Vukovic et al., 2013). It is plausible that the worry associated with maths anxiety and anticipation of performance within the classroom, may cause a distraction for the maths anxious learner. It is also possible that for the less anxious student, a certain amount of maths anxiety may be a stimulating environment, which encourages the learner to do well.

Performance itself (and additional mediating processes) is an important factor in future performance. For example, successful performance has been shown to lead to more practice and consequently, better performance (Jansen et al., 2013). As maths anxious individuals are more likely to avoid mathematical activities, a reciprocal relationship may exist between maths anxiety and performance, where performance drops as a consequence of anxiety leading to more avoidance and less opportunity for future success (Wu, Amin, Barth, Malcarne, & Menon, 2012). One study which explored the causal ordering between maths anxiety and performance in a sample of US students from 7th through to 12th grade, posited three alternative models: 1) high maths anxiety leads to poor maths performance; 2) poor maths performance leads to high levels of maths anxiety; and 3) there is a reciprocal relationship between the two constructs (Ma & Xu, 2004). Using panel analysis, the study found that lower maths achievement in the early grades led to higher maths anxiety in later grades. Some evidence was found for higher maths anxiety leading to later lower maths achievement, but the effect was very weak and only found in the early grades. The study concluded that overall, the direction of effects was from achievement to anxiety rather than vice versa or reciprocal. The study also found different patterns across gender whereby females' prior low maths achievement predicted later high maths anxiety only at transition points (e.g. from elementary to junior high school and from junior high to senior school). Boys' prior maths achievement consistently predicted maths anxiety throughout the six grades tested.

While there have been several studies investigating associations between maths anxiety and maths performance longitudinally, to date there is limited research taking both a longitudinal and a cross-cultural approach. Much of the cross-cultural focus has consisted of cross-sectional comparisons between Western and Asian cultures (e.g. Lee, 2009). One study which comprised a 'Russian' sample, compared several European countries that included Latvia with Confucian Asian countries (Morony, Kleitman, Lee, & Stankov, 2013). The Latvian sample was large enough to be divided into two ethnic and linguistic groups – Latvian and Russian. Latvia ranked in similar positions to Russia and the UK in the 2011-12 Programme for International Student Assessment (PISA; Woessmann, 2016). The study reported that associations between maths anxiety and maths accuracy were similar across all the participating countries (average r = -.27). This figure, is slightly below the average correlations (r = -.3) reported previously in the literature.

Little research is available on the relationship between academic anxiety and performance beyond the domain of mathematics. An association has been shown between second language learning anxiety and second language achievement (Horwitz, 2001). However, a literature search showed no research into geography anxiety. Geography was chosen as a comparative subject for this thesis as while it also contains mathematical and spatial content, it is taught and perceived differently to maths. Research has shown evidence of spatial anxiety, however, in relation to navigation efficiency. The study showed a positive association (r = .3) between spatial anxiety and navigational errors (Hund, & Minarik, 2006).

A recent study investigated measures of spatial anxiety and found a moderate correlation between spatial anxiety (two measures: rotational/visualization anxiety; and navigational anxiety) and mathematical anxiety (r = .32 and .41, respectively). This correlation was partly explained by the same genetic effects (navigational anxiety: 38%; and rotational/visualization anxiety: 41%) (Malanchini et al., 2017). Given the link between spatial ability and maths achievement (Wai, Lubinski, & Benbow, 2009) it is unsurprising that a moderate association and shared aetiology should be found between maths anxiety and spatial anxiety. With this in mind and the shared spatial and numerical content between maths and geography subjects, it is plausible to expect to see an association between geography anxiety and geography performance, similar to that seen between maths anxiety and maths performance.

Studies have demonstrated reciprocal associations between academic anxiety for some specific school subjects and performance in these subjects. These associations remain unexplored for geography anxiety and performance. Using longitudinal modeling, this study investigates the development of associations across one academic year between maths anxiety and maths performance, as well as between geography anxiety and geography performance. Analyses reported in Chapter 5 of this thesis showed an effect of classroom for maths anxiety in the UK sample but not in the Russian sample. With this in mind, it is expected that associations between anxiety and performance will also differ between the two samples. In the UK, where students are streamed by ability for their maths classes, they are tested at the end of each half term (every six weeks), across the academic year. These test results predict whether they move classes (to a higher or lower ability level) or remain in the same one. While this process enables students to move up if their grades improve, this likely exerts pressure on students to perform well at the tests. Moving classes is a positive experience for some students and a negative one for others. Some students will likely enjoy the challenge of the higher ability classroom. Other students may want to remain with their peers and not move up or down. Some may not want to move back to a lower ability class if they only recently moved up. Consequently, it might be expected to see a stronger association between maths anxiety and performance in the UK than in Russia.

As students are not streamed in the UK for geography, and there is less emphasis on students to achieve well at this subject relative to maths, weaker associations would be expected between geography anxiety and geography performance. Similarly, in the Russian sample, with the emphasis to do well at maths, stronger patterns of association would be expected between maths anxiety and maths performance compared to associations between geography anxiety and performance.

The Current Study

The data were collected at three assessment waves over the course of one academic year in four urban schools, two in the UK and two in Russia. All schools are mixed ability, although in the UK, students are streamed by ability for their maths classes. Based on previous research and results reported in Chapter 5 of this thesis, the current study addresses the following research questions: 1) Do associations between maths anxiety and maths performance develop differently for students in Russia and the UK? For example, in the UK sample, where students were streamed for maths, are associations between maths anxiety and performance stronger than in the Russian sample? 2) Do reciprocal associations exist between geography anxiety and geography performance, as previously shown for mathematics and other academic domains? 3) Do associations between geography anxiety and geography performance develop differently for students in Russia and the UK?

Methods

Participants

Participants were 520 10 to 12 year old students, from four urban mixed ability schools; two in London, UK and two in St. Petersburg, Russia. Although the UK schools were mixed ability, students were streamed by ability for their maths classes. The Russian students were not streamed for ability. However, they attended schools with specialized linguistic programmes. All students were in the first year of their secondary education, with specific subject teachers for the first time. Students with special educational needs were excluded from these analyses. A more detailed description of the sample used in this study is provided in Chapter 3, page 84.

Measures

A detailed description of the measures used in this study is provided in the methods section in Chapter 2, pages 61 to 70.

Procedure

A detailed description of the procedure used in this study is provided in Chapter 3, page 85.

Analyses

Analyses were conducted using data collected from the UK schools at time 2, time 3 and time 5 to correspond with the data collections in Russia at time 1, time 2, and time 3 on variables with outliers (±3SD) removed. In order to remove any effects of differences in age, age was used as a covariate. Prior to the main analyses, bivariate correlations assessed the stability of the constructs within each sample across the three assessment points.

Cross-lagged panel analysis was used to investigate longitudinal associations between performance, and academic anxiety. The cross-lagged design allows the examination of the stability of (or changes in) individual differences in a trait across time (autoregressive effects). It also shows cross-sectional links between two or more traits. This enables investigation of cross-lagged effects i.e., relationships between temporally preceding variables which stem from excess variability not explained by autoregressive and cross-sectional effects (Geiser, 2013). For example, in terms of maths anxiety and maths performance an autoregressive effect is obtained by regressing maths anxiety at time 2 on maths anxiety at time 1. This indicates how maths anxiety at time 1 predicts maths anxiety at time 2. The same is repeated for maths performance at time 1 and time 2. The cross-sectional effect is obtained by correlating maths anxiety at time 1 with maths performance at time 1, and repeating this for time 2. The cross-lagged effect between these two variables is the path between maths anxiety at time 1 and maths performance at time 2, and/or maths performance at time 1 and maths anxiety at time 2. In these analyses where three assessment points are used, the basic model is extended to include the third assessment point variables. The models presented here estimated first-order autoregressive effects which are between adjacent assessments, and second-order effects which are between the first and third assessments.

These analyses used multi-group methodology which enables testing of equivalence of parameters across more than one group or sample (Jöreskog, 1971). In this case, it was used to make comparisons of path coefficients between models for the UK and Russian samples. Analyses began with a baseline model, where all the parameters were freely estimated. Following this, three incrementally more restrictive models were tested where specific parameters were constrained across the groups: Model 1) Baseline model but with equal autoregressive effects; Model 2) Model 1 plus equal cross-lagged effects; Model 3) Model 2 with equal error variance. Each model was estimated using the maximum likelihood method which gave a large sample χ^2 goodness of fit test for the overall model (Jöreskog, 1971). The best fitting model (non-significant χ^2)indicated the stage of measurement restriction at which the groups were equal. For example, if model three was the best fit, then autoregressive, cross-sectional and cross-lag paths were equivalent across groups, i.e. not significantly different (Byrne, 2012). The Mplus 7.4 software package was used to run the analyses (Muthen & Muthen, 2015).

Results

For descriptives of the assessed variables please refer to Appendix 1, Tables 1.6 and 1.7.

Correlations Between Variables

Bivariate correlations were estimated separately for each sample and separately for maths and geography variables across the three assessment waves.

Maths anxiety and maths performance

Correlations between maths performance and maths anxiety at times 1, 2 and 3 are presented in Table 6.1 (UK) and 6.2 (Russia). Moderate to strong correlations were shown for individual constructs across all assessment waves indicating their stability across the academic year within both countries. For the

UK, the associations were slightly stronger (maths anxiety average r = .57 maths performance average r = .76), compared to Russia (maths anxiety average r = .54; maths performance average r = .58). Weak to moderate negative correlations were revealed between the two measures (UK: r = -.16 to -.37; Russia: r = -.15 to -.23). Age, included because it was used as a covariate in the cross-lagged model, associated with maths performance at time 1 for the Russian sample only.

Table 6.1. Bivariate correlations between maths anxiety, maths performance at time 1, time 2 and time 3, and age (N) for the UK sample

	Maths PVT T1	Math anx T1	Maths PVT T2	Maths anx T2	Maths PVT T3	Maths anx T3
Maths PVT T1	1 (291)					
Maths anx T1	-.26** (288)	1 (289)				
Maths PVT T2	.75** (268)	-.24** (265)	1 (286)			
Maths anx T2	-.19** (264)	.54** (262)	-.26** (282)	1 (282)		
Maths PVT T3	.79** (275)	-.37** (272)	.75** (266)	-.29** (262)	1 (294)	
Maths anx T3	-.24** (270)	.57** (267)	-.16** (262)	.60** (258)	-.26** (289)	1 (289)
Age	.01 (290)	.02 (288)	-.10 (267)	-.02 (263)	-.01 (274)	-.02 (269)

Maths PVT = maths performance. Scale: 1-48; **Maths anx** = maths anxiety, scale 1-5 where 5 = high anxiety; **T** = time; **. Correlation is significant at the 0.01 level (2-tailed) *. Correlation is significant at the 0.05 level (2-tailed).

Table 6.2. Bivariate correlations between maths anxiety, maths performance at time 1, time 2 and time 3, and age (N) for the Russian sample

	Maths PVT T1	Math anx T1	Maths PVT T2	Maths anx T2	Maths PVT T3	Maths anx T3
Maths PVT T1	1 (229)					
Maths anx T1	-.17* (220)	1 (220)				
Maths PVT T2	.59** (204)	-.18* (197)	1 (222)			
Maths anx T2	-.21** (194)	.55** (188)	-.19** (211)	1 (212)		
Maths PVT T3	.55** (191)	-.14 (185)	.62** (191)	-.13 (182)	1 (220)	
Maths anx T3	-.22** (189)	.46** (184)	-.23** (190)	.63** (181)	-.15* (213)	1 (219)
Age	-.16* (229)	-.01 (220)	-.13 (204)	-.07 (194)	-.12 (191)	.00 (189)

Maths PVT = maths performance. Scale: 1-48; **Maths anx** = maths anxiety, scale 1-5 where 5 = high anxiety; **T** = time; **. Correlation is significant at the 0.01 level (2-tailed) *. Correlation is significant at the 0.05 level (2-tailed).

Geography anxiety and geography performance

Correlations between geography performance and geography anxiety at times 1, 2 and 3 are presented in Table 6.3 (UK) and 6.4 (Russia). Moderate to strong correlations were shown for individual constructs across all assessment waves indicating their stability across the academic year within both countries. Similarly to maths, the associations for the UK were slightly stronger (geography anxiety average r = .59; geography performance average r = .59), compared to Russia (geography anxiety average r = .51; geography performance average r = .59). Weak negative correlations were revealed between the two measures in the UK across the three waves (r = -.15 to -.26), and in Russia only between geography performance at time 1 and geography anxiety at time 3 (r = -.18). Age, included because it was used as a covariate in the cross-lagged model, showed no significant associations for both countries.

Table 6.3. Bivariate correlations between geography anxiety, geography performance at time 1, time 2 and time 3, and age (N) for the UK sample

	Geog PVT T1	Geog anx T1	Geog PVTT2	Geog anx T2	Geog PVT T3	Geog anx T3
Geog	1					
PVT T1	(287)					
Geog	-.20**	1				
anx T1	(260)	(265)				
Geog	.56**	-.18**	1			
PVT T2	(264)	(243)	(286)			
Geog	-.15*	.59**	-.25**	1		
anx T2	(248)	(233)	(268)	(268)		
Geog	.63**	-.20**	.58**	-.16*	1	
PVT T3	(269)	(248)	(265)	(248)	(293)	
Geog	-.20**	.57**	-.19**	.60**	-.26**	1
anx T3	(260)	(240)	(257)	(241)	(281)	(282)
Age	-.04	.07	-.01	.00	.02	.05
	(286)	(264)	(267)	(250)	(273)	(264)

Geog PVT = geography performance, scale: 1-37; **Geog anx** = geography anxiety, scale: 1-5 where 5 = high anxiety; **T** = time; **. Correlation is significant at the 0.01 level (2-tailed) *. Correlation is significant at the 0.05 level (2-tailed).

Table 6.4. Bivariate correlations between geography anxiety, geography performance at time 1, time 2 and time 3, and age (N) for the Russian sample

	Geog PVT T1	Geog anx T1	Geog PVT T2	Geog anx T2	Geog PVT T3	Geog anx T3
Geog	1					
PVT T1	(227)					
Geog	-.13	1				
anx T1	(205)	(207)				
Geog	.59**	-.10	1			
PVT T2	(202)	(184)	(220)			
Geog	.03	.45**	-.03	1		
anx T2	(197)	(180)	(213)	(217)		
Geog	.53**	-.04	.64**	.06	1	
PVT T3	(192)	(175)	(192)	(188)	(224)	
Geog	-.18*	.47**	-.12	.60**	-.03	1
anx T3	(188)	(174)	(188)	(186)	(218)	(221)
Age	-.02	.06	.07	-.02	-.01	-.07
	(227)	(207)	(203)	(199)	(193)	(190)

Geog PVT = geography performance, scale: 1-37; **Geog anx** = geography anxiety, scale: 1-5 where 5 = high anxiety; **T** = time; **. Correlation is significant at the 0.01 level (2-tailed) *. Correlation is significant at the 0.05 level (2-tailed).

Cross-Lagged Links Between Maths Anxiety And Maths Performance

These analyses were conducted to assess whether associations between maths anxiety and maths performance develop differently for students in Russia and the UK. The model in Figure 6.1 shows the longitudinal associations between maths anxiety and maths performance across one academic year at times 1, 2 and 3 with age used as a covariate. The significant cross-lagged paths are presented separately for the UK and Russia. The model presented is the baseline model which, having satisfactory fit, provided the best fit to the data, AIC = 6926.69; BIC =7190.67; $\chi^2(520)= 14.625$, $p = 0.01$; RMSEA = 0.10; CFI = 0.99; TLI = 0.90; SRMR = 0.02. The subsequent models constraining path coefficients to be equal between the two countries showed considerably worse fit, indicating that the emergent associations between maths anxiety and maths performance were significantly different between the two countries. This is particularly evident for the cross-lagged effects between the two countries. In the UK, maths anxiety at time 2 negatively influenced maths performance at time 3; whereas in Russia, maths performance at time 1 negatively influenced maths anxiety at time 2. Age also negatively associated with maths performance at time 1 for Russia. The model is described below for each sample:

UK. The model shows that all autoregressive effects (including first- and second-order) were significant. This represents that a significant amount of individual differences remained stable across time for both maths anxiety and maths performance. Less stability was observed between time 2 and 3, for maths performance, which was accounted for by the significant cross-lagged path from maths anxiety at time 2 to maths performance at time 3. The negative relationship indicates that higher maths anxiety associated with subsequent lower maths performance. This association accounted for a small amount of variance in the model, as the cross-lagged coefficient was small compared to the autoregressive effects. The slightly larger cross-sectional effects shown at time 1 and 2 indicate a modest amount of shared variance between maths anxiety and maths performance at these assessment points over and above that explained by the autoregressive effects. These associations were possibly due to situation-specific effects that impacted both measures at the same assessment point. This effect was not maintained at time 3. Perhaps students were maths anxious when completing the maths performance task during data collection at time 1 and 2, but not at time 3 as they had become used to the assessment. The sum of the autoregressive and cross-lagged effects in the model explained 56% of the variance in maths performance and 27% of the variance in maths anxiety at time 2 (R^2 = 0.56 and 0.29 respectively). At time 3 the amount of variance explained was 70% (R^2= 0.70) for maths performance and 43% (R^2= 0.43) for maths anxiety.

Russia. Similarly to the UK, the model shows that all (first- and second-order) autoregressive effects were significant, suggesting that for Russia a significant amount of individual differences remained stable across time for both maths anxiety and maths performance. Slightly less stability was observed for maths anxiety between time 1 and 2, compared to between time 2 and 3, which was accounted for by the significant cross-lagged path between maths performance at time 1 and maths anxiety at time 2. The negative relationship suggests that high maths performance associated with subsequent lower maths anxiety. This association accounted for a small amount of variance in the model, as the cross-lagged coefficient was small compared to the autoregressive effects. Age also negatively associated with maths performance at time 1. This suggests that age also contributed towards the relationship between maths performance at time 1 and maths anxiety at time 2. The significant negative cross-sectional effect observed at time 1 suggests a modest amount of shared variance between maths anxiety and maths performance at time 1. As this association was not maintained across subsequent assessments, it suggests that situation-specific effects impacted at time 1 only. It may be that students experienced maths anxiety when completing the maths performance during data collection at time 1, but not at subsequent assessment waves. The amount of variance explained by the sum of the autoregressive and cross-lagged effects in the model indicates that 34% of the variance in maths performance and 33% of the variance in maths anxiety is explained at time 2 (R^2 = 0.34 and 0.33 respectively). At time 3 the amount of variance explained is 44% (R^2= 0.44) for maths performance and 48% (R^2= 0.48) for maths anxiety.

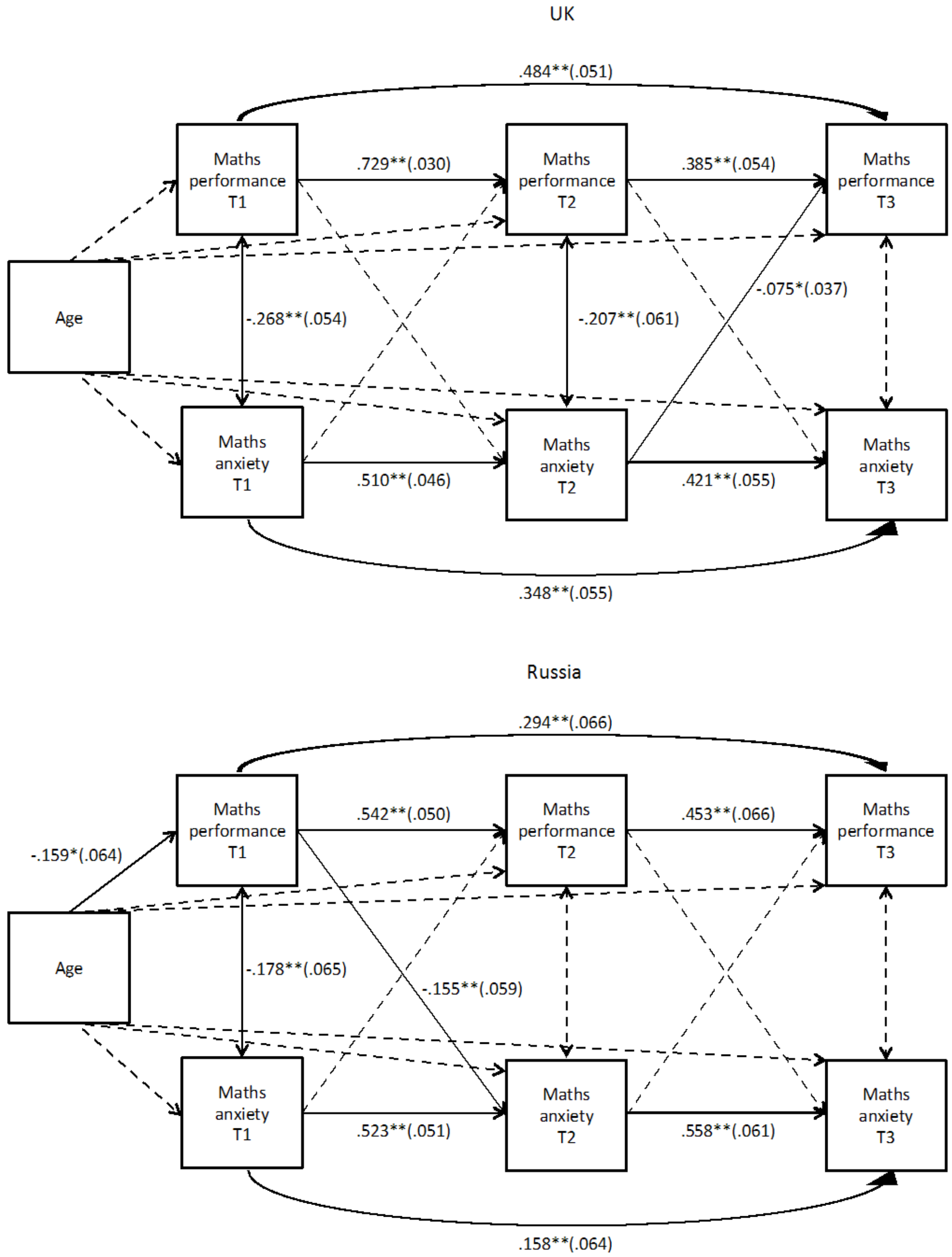

Figure 6.1. Multi-group cross-lagged analyses comparing the UK and Russia (top and bottom, respectively) for the relationship between maths anxiety and maths performance across three waves of assessment during one academic year (time 1, time 2 and time 3) with age as a covariate. Standardised coefficients are presented with standard errors in parenthesis. T=time; significance is indicated by *$p \leq .05$, ** $p \leq .01$.

Cross-Lagged Links Between Geography Anxiety And Geography Performance

Further analyses were conducted to explore whether reciprocal associations exist between geography anxiety and geography performance, as previously shown for mathematics, and whether they develop differently for students in Russia and the UK. The model in Figure 6.2 shows the longitudinal associations between geography anxiety and geography performance across one academic year at times 1, 2 and 3 with age used as a covariate. The significant path coefficients are presented separately for the UK and Russia. The model presented is the baseline model which having acceptable fit, provided the best fit to the data, AIC = 6872.64; BIC =7136.38; $\chi^2(520)$= 11.965, $p = 0.02$; RMSEA = 0.09; CFI = 0.99; TLI = 0.90; SRMR = 0.02. The subsequent models, constraining path coefficients to be equal between the two countries, showed considerably worse fit, indicating that the emergent associations between geography anxiety and geography performance were significantly different between the two countries. This is particularly evident when examining the cross-sectional effects between the two countries. In the UK, cross-sectional effects persisted across each assessment point, whereas in Russia no cross-sectional effects were observed. No cross-lagged paths were significant in either sample. The model is described below for each sample:

UK. The model shows that all autoregressive effects (including first- and second-order) were significant. This represents that a significant amount of individual differences remained stable across time for both geography anxiety and geography performance. It also indicates that subsequent performance was predicted by earlier performance and subsequent anxiety was predicted by earlier anxiety across the model. Slightly less stability was observed between time 2 and 3, compared to between time 1 and 2, for both geography anxiety and geography performance. The cross-sectional effects shown at time 1, 2 and 3 indicate a modest amount of shared variance between geography anxiety and geography performance at these assessment points over and above that explained by the autoregressive effects. These associations were possibly due to situation-specific effects that impacted both measures at the same assessment point; effects that persisted across each wave. For example, students may have experienced geography anxiety when completing the geography performance task during data collections at each assessment wave. The amount of variance explained by the sum of the autoregressive effects in the model indicates that 31% of the variance in geography performance and 36% of the variance in geography anxiety is explained at time 2 (R^2 = 0.31 and 0.36 respectively). At time 3 the amount of variance explained is 47% (R^2= 0.47) for geography performance and 45% (R^2= 0.45) for geography anxiety.

Russia. Similarly to the UK, the model shows that all (first- and second-order) autoregressive effects were significant, suggesting that for Russia a significant amount of individual differences remained stable across time for both geography anxiety and geography performance. This also indicates that subsequent performance was predicted by earlier performance and subsequent anxiety was predicted by earlier anxiety across the model. Unlike the UK, there were no significant cross-sectional effects observed which suggests no situation-specific shared variance between geography anxiety and geography performance over and above that explained by the auto-regressive effects. The sum of the autoregressive effects in the model explained 34% of the variance in maths performance and 21% of the variance in maths anxiety at time 2 (R^2 = 0.34 and 0.21 respectively). At time 3 the amount of variance explained is 44%

(R^2= 0.44) for maths performance and 48% (R^2= 0.48) for maths anxiety.

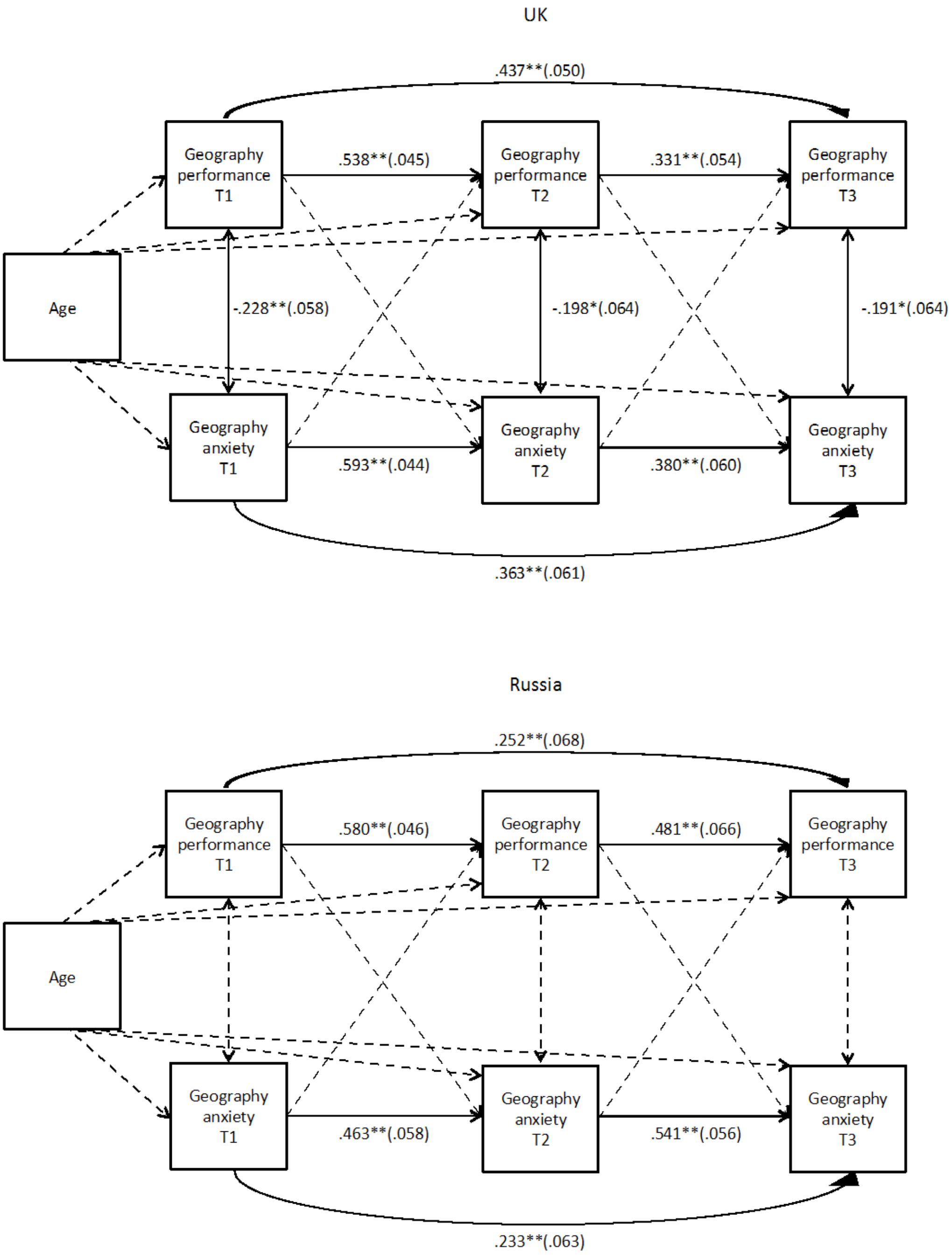

Figure 6.2. Multi-group cross-lagged analyses comparing the UK and Russia (top and bottom, respectively) for the relationship between geography anxiety and geography performance across three waves of assessment during one academic year (time 1, time 2 and time 3) with age as a covariate. Standardised coefficients are presented with standard errors in parenthesis. T=time; significance is indicated by *$p \leq .05$, ** $p \leq .01$.

Discussion

The aim of this study was to explore whether associations between maths anxiety and maths performance develop differently for students in Russia and the UK. For example, it might be expected in the UK sample, where students were streamed for maths, that associations between maths anxiety and performance would be stronger than in the Russian sample. The results from the bivariate correlations across the three assessment waves (see Tables 6.1 and 6.2) show similarity between the samples regarding stability of the constructs as moderate to strong within-construct associations were observed for both maths anxiety and maths performance, across the UK and Russian samples. Slightly stronger associations were observed, however, between maths anxiety and maths performance in the UK sample (r =-.164 to -.366), compared with the Russian sample (r = -.148 to -.234).

The results from the cross-lagged analyses where the baseline models were the best fit for the data between the two countries, also suggests that associations among the two maths constructs developed differently between the UK and Russian samples. There were some similarities across both countries, which were also evident from the bivariate correlations (Tables 6.1 and 6.2). Stability was observed for the two maths constructs with earlier maths anxiety predicting subsequent maths anxiety, and prior performance predicting later performance (see Figures 6.1). At time 1, shared situation-specific effects were observed between maths anxiety and maths performance across both countries. At time 3, for both countries, there were no such effects. However, at time 2 differences were observed as shared situation-specific effects remained for the UK only. Another difference was the direction of effects between maths anxiety and maths performance. In the UK sample, high maths anxiety at time 2 led to lower subsequent maths performance at time 3. This was contrary to previous research that suggested a causal relationship whereby prior performance predicted later maths anxiety (Ma &Xu, 2004). The results for the Russian sample, however did support this study with poor/successful performance at time 1 predicting later high/low maths anxiety at time 2. As age also associated with maths performance at time 1 for the Russian sample, this suggests that age may be a factor that influenced differences in performance at time 1 which in turn, influenced differences in later maths anxiety. The difference in causal ordering between the two samples is interesting. The finding suggests, contrary to Ma and Xu's conclusion, that more than one model likely exists between maths performance and maths anxiety. In the present study, the focus was on the development of associations between maths anxiety and performance within one academic year. It is possible that by using several assessments within one year the present study was more able to detect incrementally, the dynamic nature of associations between maths anxiety and performance, whereas Ma and Xu's study evaluated the associations more broadly with yearly assessments across six grades. Ma and Xu did find evidence for earlier high maths anxiety associating with subsequent low performance in students close in age to the present sample at grades 7 and 8 (ages 12 to 13 years). The effects however, were weak compared to the alternative model. Consequently, Ma and Xu concluded that overall, earlier achievement predicted later maths anxiety. Their finding of different patterns across gender whereby prior low maths achievement predicted females' later high maths anxiety only at transition points (e.g. from elementary to junior high school and from junior high to senior school), suggests some variation even in their model. It is possible that for the present study, more variation in effects were likely to emerge because the timing was around transition from primary to secondary education. A time when the curriculum becomes more intensive, and other factors, such as maturation processes are in influence (Eccles, 1999).

The difference in causal-ordering of maths anxiety and maths performance between the two samples may be due in part to the differences between them in terms of streaming processes. It may be that in a non-streamed environment, as in the Russian sample, anxiety is predicted by performance, as performance is compared with performance of weaker and stronger peers. Whereas in a streamed environment, as in the UK, the comparison of performance may be attenuated because students are more similar to each other in ability streamed classes. Therefore, other factors, such as individual differences in anxiety may explain more variance in performance. It may also be that in the UK sample where a stringent streaming process was employed in the samples' schools it may have intensified the relationship between the two constructs. Indeed, slightly stronger associations were evident between maths anxiety and maths performance in the UK compared to the Russian sample. It may be that the streaming process exerted a considerable amount of unhealthy pressure on students, made worse perhaps if students were prone to higher levels of maths anxiety. By time 2 and 3 (where the effect was shown), when the students had gone through the 6-weekly testing and streaming process four times, their levels of anxiety may have been heightened to the degree that it impacted their later performance.

Another aim of the study was to investigate whether longitudinal associations develop between geography anxiety and geography performance, in a similar way to those observed in the literature between maths anxiety and maths performance. The results from the bivariate correlations across the three assessment waves (Tables 6.3 and 6.4) suggest that similarly to maths, geography constructs were also stable with moderate to strong within-construct associations observed for both geography anxiety and geography performance. Slightly weaker associations were observed between geography anxiety and performance compared to associations between maths anxiety and maths performance.

The results from the cross-lagged analyses suggest that associations between geography anxiety and performance develop differently to those shown between maths anxiety and maths performance. More specifically, no causal relationships between geography anxiety and geography performance were found over time. While geography anxiety existed in these samples, it did not predict later performance. Equally, geography performance did not predict later geography anxiety.

The difference in between-construct associations for maths and geography domains may be related to the testing situation during the assessments. Maths anxiety has been shown to disrupt maths performance more acutely in a testing situation (e.g. Ashcraft & Moore, 2009). In this study the maths performance task used was a timed test which likely applied more pressure on the students during data collection. It was not possible to time the geography task because with the larger amount of reading content, individual differences in reading ability would likely confound test results. The combined pressure on students to perform well at maths generally and the timed nature of the test, may have led students to feel more anxious during the maths performance task, despite reassurances that their results would not be ‘judged’ and confidentiality of their responses would be strictly maintained.

Another potential reason for the difference between the results for mathematics and geography is the level of importance associated with each domain. For example, there is a huge emphasis for students to perform well at maths, especially as future career choices depend on good grades. There is little emphasis however on geography achievement, despite its importance in educating students, not only in navigational skills but also about the world, about similarities and differences of populations, climate and environmental issues, to name just a few.

Another aim of the study was to investigate whether associations between geography anxiety and geography performance develop differently between Russia and the UK. The absence of causal relationships between geography constructs is similar across the UK and Russian samples. However, as the baseline models were the best fit for the data between the two countries, it suggests that associations among the two geography constructs developed differently between the UK and Russian samples. There were some similarities, however, as shown in the correlations, stability was observed for the two geography constructs with earlier geography anxiety predicting subsequent geography anxiety, and prior performance predicting later performance (see Figures 6.2). Shared situation-specific effects were observed at each assessment wave for the UK only, suggesting that only UK students experienced geography anxiety when completing the geography performance task during each data collection. As no such associations were observed for the Russian sample, it suggests no relation between geography anxiety and geography performance in this sample.

The difference across the two samples in associations between geography anxiety and performance may be due to the mathematical content in geography. The UK sample showed a slightly stronger relationship between maths anxiety and maths performance than the Russian sample. They may, therefore, have been more acutely affected by the mathematical content in geography. The difference may also stem from differences in curricula between the two countries. In the UK sample, geography was studied as part of a humanities course comprised of several subjects which alternated across each semester. The UK students may have perceived themselves as less confident in geography compared to the Russian sample, who studied the subject throughout the academic year. This may have increased the UK sample's anxiety levels during the assessment, despite their slightly better performance found in Chapter 3. These findings highlight the cross-cultural differences in academic anxiety.

Limitations and Future Directions

The study is not without limitations. As mentioned in Chapters 3, 4 and 5 of the present thesis, the timing of the first data collection for the Russian sample was one issue. The first data collection occurred at the beginning of the spring term rather than the beginning of the academic year. This meant that when data were matched to corresponding data collections in the UK, an earlier assessment point for the UK was missed. Although with the short time lags, this was unlikely to be a huge issue, it would have been interesting to observe potential effects at the very start of the academic year. Furthermore, while the study was able to investigate the dynamic nature of the relationships across one academic year, it may also have benefitted from further data collections in subsequent years to explore more broadly across years of education.

One limitation concerns the models, while the fit indices suggested acceptable fit, these were not optimum fitting models. One potential reason for inadequate fit is the strength of correlations between the anxiety constructs and performance constructs. For maths, the association between anxiety and performance was weak to moderate, and slightly weaker for the Russian sample compared to the UK. For geography, any association between anxiety and performance was weak and only found in the UK sample.

Another limitation is that the study was unable to compare the countries on school achievement as these grades were not comparable between the two samples. This would have been an interesting addition to the study.

Future investigations will explore potential relationships with maths anxiety for different aspects of maths, for example number estimation. As previous research has shown an association between maths anxiety and maths self-efficacy (e.g. Cooper & Robinson, 1991; Goetz, Bieg, Lüdtke, Pekrun, & Hall, 2013), it would be interesting to explore these associations within and between the two samples. Further, investigations will also consider teacher-student relations, particularly as differences between classrooms and teacher groups were found in the studies reported in Chapters 4 and 5. Investigations will also be explored independently for the samples to take account of additional waves of assessment in the UK.

Conclusion

The study showed that longitudinal associations between maths anxiety and maths performance developed differently across one academic year between the Russian and UK samples. The between-construct differences in the strength of associations and the between-sample differences in causal ordering indicate the complexity of the relationship between maths anxiety and performance, which likely depends on other factors such as streaming. Cross-domain disparity such as the absence of causal relationships between geography anxiety and performance in both samples, may result from the different implementation of the maths and geography performance tasks during data collections. They may also be due to unequal levels of importance associated with these two academic subjects. Taken together with the dissimilarity across samples for associations between geography anxiety and performance, this study shows that longitudinal associations between academic anxiety and academic performance manifested differently cross-culturally, and developed differently between academic subjects. Variation found cross-culturally may be a consequence of dissimilarities in education systems.

Chapter 7

Twin classroom dilemma: to study together or separately?

Abstract

There is little research to date on the academic implications of teaching twins together or separately. Consequently, it is not clear whether twin separation in educational contexts leads to positive or negative outcomes. As a result, parents and teachers have insufficient evidence to make a well-informed decision when twins start school. This study addresses this issue in two large representative samples of twins from Quebec (Canada) and the UK. Twin pairs taught together and taught separately were evaluated across a large age range (7 to 16 years) on academic achievement, a range of cognitive abilities and motivational measures. Overall, results showed no average positive or negative effects of classroom separation on children's academic achievement, cognitive ability and motivation. The results are discussed in terms of cultural and educational similarities and differences across Quebec and the UK, and suggest guidelines for policymakers. (See graphical abstract in Appendix 6, Figure 6.1).

Note: This chapter is being submitted as a multi-author publication. As joint co-author I conducted all analyses for the UK sample and co-wrote the manuscript and the supplementary materials in Appendix 6.

Introduction

The twin and multiple birth association (TAMBA) in the United Kingdom (UK), recommend that the decision of whether to educate twin pairs separately or together should be one made by parents and teachers (TAMBA, 2009; 2010). Separation might have positive consequences: aiding development of individual identities, reducing inter-twin competition (Segal & Russell, 1992), and decreasing dependency, especially where dominant-dominated relationships occur (Lalonde & Moisan, 2003). For practical reasons, separating twins helps teachers and other class members to distinguish between the pair.

Conversely, the arguments against separation are also strong. A recent study found that twinship may have a positive effect on longevity, similar to a documented positive effect of marriage on longevity (Sharrow & Anderson, 2016). It is possible that the protective effect of twinship results from the unique bond held between twin pairs. Considering the proximity that twins have shared all their lives up to the beginning of school, separation from their co-twin at this point may have adverse emotional consequences (e.g. Van Leeuwen, Van Den Berg, van Beijsterveldt, & Boomsma, 2005; Tully, Moffitt, Caspi, Taylor, Kiernan, & Andreou, 2004). Further, separation anxiety, co-twin preoccupation, and increased desire to be with their co-twin may reduce school enjoyment (Lalonde & Moisan, 2003). On a practical level, if the twins attend different schools, getting both twins to school on time presents a logistical problem for parents, which might contribute to family stress.

Choice for separation may reflect twins' interests/suitability for a specialised school or program (e.g. specialised music school or schools with enhanced maths curricula). It may also reflect imposed selection processes of setting or streaming to different schools/classes by ability. For twins, this may result in a higher number of separated non-identical twins (dizygotic; DZ) compared to identical twins (monozygotic; MZ) as MZ twins are more similar in ability (Petrill, Kovas, Hart, Thompson, & Plomin, 2009) and motivation (Spinath, Spinath, & Plomin, 2008).

Considering the lack of solid empirical evidence regarding positive and negative outcomes of twin separation, it is unsurprising that different countries have different policies regarding educating twins together or separately.

Policies Around the World

In Quebec (Canada), separation of twin pairs is widespread. Canada's policy for classroom placement of multiple births is to leave the decision to parents, although separating twins is sometimes strongly encouraged by the School Commission Boards (Lalonde & Moisan, 2003). Similarly, in the Netherlands, the advisory board Dutch Society of Parents of Multiples (NVOM) advocates separation (Van Leeuwen et al., 2005), although parents are advised to make their decision on what they believe is best for their children. In the UK, parents can mostly choose whether or not to send twins to the same class. A recent survey of 514 UK parents of twin pairs aged up to 3 years showed that 60% of MZ and 55% of DZ twins' parents wanted to keep the twins together when they start school (Cherkas, 2015). However, in around 20% of cases, schools have a stringent policy to separate twins and/or triplets without consultation or supporting evidence that this would be in the children's best interests (Cherkas, 2015). In the USA, twin's classroom separation decision is left to the school. However, parental opinion about classroom separation is divided. US mothers of young twins and triplets (n=63) were surveyed with respect to their satisfaction regarding schools' decision to separate or not to separate twins. Approximately half of the parents did not support a general practice of separating twins, while the other half was in favour of classroom separation (Segal & Russell, 1992). For Russian twins, while there is no clear policy, existing practice favours non-separation, unless twins are attending separate specialized schools. In contrast, most twins are educated separately in China. Data from a sample of 820 Chinese twin pairs, as part of The Beijing Twin Study (BeTwiSt: Chen, Li, Zhang, Natsuaki, Leve, Harold, & Ge, 2013), showed that 255 pairs were in the same class (31.1%); 432 pairs were in the same schools, but different classes (52.7%); and the other 133 pairs were in different schools (16.2%). The reasons for favouring separation in China are diverse: twins themselves and their parents might decide to be separated; some parents might send one twin to their grand-parents' home to reduce the parent's family burden; one twin may attend a good school by passing tests which his/her co-twin failed; or parents may encourage their twins to develop interpersonal relationships with other children by separating them into different classes.

Only a handful of studies to date have been conducted on cognitive and academic outcomes of twins educated separately or together, and these studies reported mixed findings. Therefore, policies and preferences regarding twin separation are not evidence-based.

Classroom Separation: Evidence From Previous Studies

A summary of previous studies investigating the effect of classroom separation on twins is given in Table 7.1. One study of Australian and US twins found no significant differences in literacy across kindergarten and 1st grade after pre-existing differences in disruptive behaviour and pre-literacy ability were taken into account (Coventry, Byrne, Coleman, Olson, Corley, Willcutt, & Samuelsson, 2009). Similarly, another study investigating the effect of separation on twins' achievement using the Netherlands Twin Registry (NTL) found no difference between separated and non-separated twin pairs at age 12 (Polderman, Bartels, Verhulst, Huizink, Van Beijsterveldt, & Boomsma, 2010). Twins taught together or separately did not differ on an independent national academic achievement test taken at the end of primary school (CITO) (controlling for zygosity, familial socioeconomic status (SES), externalising problems at age 3 and urbanisation). However, a study from a large Netherland's educational survey collected longitudinally across Grades 2 (aged 6 years) to 8 (aged 12 years) reported significantly lower language (d = 0.02) and arithmetic (d = 0.23) scores for separated twins in early school years, especially for same-sex pairs (Webbink, Hay, & Visscher, 2007). After controlling for peer achievement, and school and familial SES, results suggested that classroom separation may have some small effect in early school years. However, there was no long-lasting effect of early separation; twins' performance in language and arithmetic was not worse when separated for 3 years than when educated together.

Similarly, a longitudinal study investigated the effects of classroom separation in UK twins at ages 5 and 7 years. Twins were divided into three groups: 1) pairs who were taught together at both ages; 2) pairs who were taught together at age 5 and separately at age 7; and 3) pairs who were separated at both ages (Tully, Moffitt, Caspi, Taylor, Kiernan, & Andreou, 2004). Both MZ and DZ twins separated at age 7 only, showed lower reading scores, with small effect sizes (see Table 7.1).

Another UK study compared mean differences between twins educated together vs. separately and found that twins educated separately were marginally more different than twins educated together. This was found for school achievement and cognitive abilities, such as verbal and non-verbal reasoning, at ages 7, 9, and 10 years (Kovas, Haworth, Dale, & Plomin, 2007). However in another study, UK twins in different classrooms were no more dissimilar in their academic motivation than twins in the same classrooms at age 9 (Kovas et al., 2015). A similar study of Australian and US twins compared twin pairs based on zygosity in the same or different classrooms from kindergarten to 2nd grade. Slightly larger mean differences were shown for twin pairs taught separately compared to those taught together across time, with larger differences shown for DZ compared to MZ twins. Lower correlations were also found for both MZ and DZ twins taught separately (Byrne, Coventry, Olson, Wadsworth, Samuelsson, Petrill, Willcutt, & Corley, 2010). The effects of separate classrooms were modest (8%), but were not due to initial differences between the pairs. The slightly smaller similarity for separated twins (than for twins educated together) may result from differences in teacher-student relationships, quality of instruction and emotional support, or peer relations (e.g. Hamre & Pianta, 2005). However, research has shown that these factors may also lead to differences in achievement for twins taught together as they each

Table 7.1. Summary of previous research investigating separation and non-separation of twin pairs in the classroom

Authors (year)	Sample size	Age of sample	Significant difference of same/different classrooms	Effect size	Analyses by zygosity
Coventry et al., (2009)	1505 individual twins/triplets (752 MZ; 752 DZ)	59-77 months	No significant difference for literacy	None	MZ vs. DZ
Polderman et al., (2010)	4006 twins (839 MZ; 1164 DZ pairs)	12 years	No significant difference for school achievement	None	Zygosity; Sex; Classroom separation by zygosity.
Webbink et al., (2007)	5756 twins (2878 pairs)	6–12 years	S>D language and arithmetic at age 6	Language: *d*=0.02 Arithmetic: *d*=0.23	None
Tully et al., (2004)	1756 twins (484 MZ; 394 DZ pairs)	5 &7 years	S>D reading D>S internalizing problems	**Internalizing problems:** Separated early: age 5: MZ (*d*=0.4); DZ (*d*=0.2) age 7: MZ (*d*=0.3); DZ (*d*=0.1) Separated late: age 5: MZ (*d*=0.4); DZ (*d*=0.3) age 7: MZ (*d*=0.4); DZ (*d*=0.2) **Reading at age 7:** Separated early: MZ (*d*=0.1); DZ (*d*=0.1) Separated late: MZ (*d*=0.2); DZ (*d*=0.1)	MZ vs DZ
Kovas et al., (2007)	11482 twins (~1910 MZ; ~3830 DZ pairs)	7, 9, &10 years	No significant difference for school achievement or cognitive ability	None	MZ vs DZ
Kovas et al., (2015)	2294 twins (~382 MZ; ~764 DZ pairs)	9 years	Non significant difference for motivation	None	MZ vs DZ
Byrne et al., (2010)	1422 twins (355 MZ; 356 DZ pairs)	54–71 months	S>D literacy	Literacy: 8%of variance explained by classroom separation status	MZ vs DZ
Asbury et al., (2008)	122 twins (61 MZ pairs)	10 years	Effect for twin pairs within same classroom in school achievement	8-15% science and maths achievement	MZ only

perceive the same classroom differently (Asbury, Almeida, Hibel, Harlaar, & Plomin, 2008).

To sum up, previous studies suggest inconsistent and very modest effects of twins' classroom separation in early school years (see Table 7.1). As a result, parents and educational policymakers are left without clear evidence for educating twins separately or together. Consequently, more research into the implications of twin separation is needed. This is particularly timely as numbers of multiple births are generally increasing as a result of a growth in the use of assisted reproductive technologies, such as in-vitro-fertilization (IVF) (Office for National Statistics, 2014).

The inconsistencies of previous research may mean that effects of classroom separation differ across different measures and samples. Previous research has also suffered from a number of limitations. First, many of the studies assessed only a maximum of 3 data points, some quite close in age. Second, few studies investigated the effect of classroom separation by twin's sex and zygosity which precluded an investigation of whether the effects of separation are stronger for specific sex/zygosity groups, e.g. MZ twins, females etc. Third, previous classroom separation studies only investigated one country, not taking into account differences in cultural and/or educational systems.

The current study

The present study sought to address these limitations by comparing school achievement, motivation and cognitive ability in twin pairs taught separately and together in two large representative twin samples in the UK and

Quebec (Canada) followed longitudinally from ages 7 to 16 years. The study addresses two main research questions: 1) Are there average differences in school achievement, cognitive ability and motivation between twin pairs taught together (i.e. by the same teacher/class) and twin pairs taught separately (i.e. by different teachers/classes)? 2) Are there any differences in separation effects, in light of differences in timing of separation, purpose of separation (e.g. streaming; policy recommendations) and twins' sex or zygosity?

Methods

Participants

The two representative samples taking part in the study are: the UK Twins Early Development Study (TEDS; Haworth, Davis, & Plomin, 2012), which provided data between ages 7 and 16 years from 8705 twin pairs (3039 MZ and 5666 DZ pairs) following exclusion of data from participants with medical issues and English spoken as a second language; and the Canadian Quebec Newborn Twin Study (QNTS; Boivin et al., 2013), which provided data from 426 twin pairs (182 MZ and 244 DZ pairs) between ages 7 and 12 years.

In both samples, participant numbers vary across measures and time of data collection. Further information about the samples is provided in Appendix 6, Sample description section.

Measures And Procedure

A broad range of achievement, cognitive and motivational measures were used across all samples. These measures are briefly summarised here, with details and the overall sample size for each twin study in Appendix 6, Tables 6.1.1 to 6.1.3.

Taught Together or Taught Separately. To determine whether twin pairs were taught together or separately, teacher contact details for each twin were used from the studies' admin data for ages 7 to 12 years (QNTS) and ages 7 to 14 years (TEDS). This gave a reliable indication of whether or not twins had the same or different teacher. For UK twins at age 16, twins self-reported retrospectively at age 18, if they were in the same class as their co-twin for English, maths and science. It is important to mention that the study is not a randomized controlled trial (RCT) of twins' classroom placement. Rather, twins' classroom allocation is likely to be the result of discussion between parents, teachers and the twins themselves.

Achievement. Across both samples and for all ages apart from age 16 (UK), school achievement data were collected by teacher report. In QNTS (Quebec-Canada), teachers assessed the twins' achievement at ages 7, 9, 10 and 12 years by answering the question: "How would you rate this child's current academic achievement (in reading, writing, mathematics, and in general)?" Rating was given on a 5-point Likert's scale ranging from 1 (near the bottom of the class) to 5 (near the top of the class).

In TEDS (UK), teachers reported children's level of achievement at ages 7, 9, 10, 12, and 14 from tests that are set and marked by the teacher according to National Curriculum (NC) guidelines. The test scores contribute towards an overall level for each subject which ranges from 1-4, 1-5, and 1-7 depending on guidelines at the time of the study (1 being the lowest level). At age 16, participants reported their own grades for externally assessed internationally recognised exams, General Certificate of Secondary Education (GCSE). These exams are taken for each specific subject at age 16, which at the time of data collection was the end of compulsory education in the UK. The exams are graded A* to G with A* being the highest. Obtaining at least grade C is necessary for many further study/career options. Data for maths, English, English language and English literature were analysed in this study. Assessment guidelines can be accessed here: https://www.gov.uk/government/consultations/gcse-subject-content-and-assessment-objectives

Cognitive Abilities. Cognitive abilities were assessed in the UK sample only. Verbal ability was evaluated at ages 7, 9, 10, 12 and 14 years using a combination of age appropriate tasks from Wechsler Intelligence Scale For Children (WISC III: Wechsler, 1992). Additional verbal tests were included at ages 9, 10 and 12 from WISC-III as a Process Instrument (WISC-III-PI: Kaplan, Fein, Kramer, Delis, Morris, 1999) (see Appendix 6, Table 6.1.2)

Non-verbal ability was also evaluated using WISC III tasks at ages 7, 10 and 12 years. Additional tests were included at age 7 from McCarthy Scales Of Children's Abilities (MCSA: McCarthy, 1972), and at age 12 from Raven's Standard Progressive Matrices (Raven, Court & Raven, 1996). Non-verbal ability was assessed at age 9 using Cognitive Abilities Test 3 (CAT3: Smith, Fernandes& Strand, 2001). At age 14, an expanded version of the age 12 Raven's Standard Progressive Matrices task was used (Raven et al., 1996).

General cognitive ability (g) was assessed using composites of the verbal and non-verbal tests for each age. Reading ability was evaluated at ages 7 and 12 using TOWRE tests of sight word efficiency and phonemic decoding efficiency (word and non-word tests) (Torgesen, Wagner, &Rashotte, 1999). At age 10, reading ability was assessed using the Peabody Individual Achievement Test (PIAT: Markwardt, 1997). Tests were adapted to be administered at age 7 by telephone. At other ages, tasks were adapted for web-based assessment. A full description of the tasks can be found in Kovas et al., (2007).

Motivational Constructs. Motivational constructs were self-reported by the children in both samples. In the QNTS, children self-reported their enjoyment, and how they perceived their ability in mathematics and reading at ages 10 and 12 with six items from the Elementary School Motivation Scale (Guay et al., 2010): 1). For example, for enjoyment: I like mathematics/reading; mathematics/reading interest me a lot; I do mathematics/reading even when I am not obliged to do so; 2) and for self-perceived ability: (SPA) mathematics/reading is easy for me; I have always done well in mathematics/reading; I learn things quickly in mathematics/reading. Children answered each item using a 4-point Likert scale ranging from 1 (never) to 4 (always). In addition, the teacher–child relationship was assessed from ages 7 to 12 through teacher rating items from the Closeness and Conflict subscales of the Teacher–Child Relationships scale (STRS; Pianta, Steinberg, & Rollins, 1995). The scale measures teachers' perception of the relationships with individual students. Scores ranged from 1 to 5, with highest scores indicating a positive relationship.

In TEDS, children completed the motivational tasks by a combination of telephone interviews and booklet completion at age 9, and by web-based testing for age 12. Children reported their enjoyment (how much do you like) and self-perceptions of ability (how good do you think you are) for solving number and money problems; doing maths in your head; and multiplying and dividing. Participants responded using a 5-point scale where 1 = very good or like very much and 5 = not good at all, and don’t like at all (Spinath, Spinath, Harlaar, & Plomin, 2006).

Although the measures were not identical across the samples, they tap into achievement and motivational constructs. As a consequence, similarity of results across the samples increases confidence in their generalizability.

Educational Policy in the UK and Quebec (Canada). UK and Quebec education systems are mostly similar with some differences in teacher/classroom allocation across the school years. In both Quebec and the UK the same teacher teaches all subjects for students during elementary/primary education, with the teacher changing on a yearly basis. In Quebec, elementary education continues to age 12 (Grade 6), whereas in the UK, primary education continues to age 11 (Year 6). In secondary education/high school, the majority of the UK schools' maths and English classes are selected based on students' ability in these subjects, while there is no such selection in Quebec, except for optional advanced classes for English.

Analyses

Analyses were conducted using one twin selected randomly from each pair, and within each sample on variables corrected for age, with outliers (±3SD) removed. Descriptive analyses assessed frequency of twins in the same vs. different classes. Chi-square analysis assessed frequency differences of groups as a function of same/different class and zygosity. Analyses of variance (ANOVA) were conducted to assess potential differences in means for achievement, cognitive ability and motivation between twins taught in the same vs. different classes by zygosity and sex by zygosity. For these analyses a Bonferroni multiple testing correction was set of $p \leq .001$ where $p = .05$ was divided by the number of measures (k = 76) across ages 7 to 16 years, across the two samples. Finally, within-pair ANOVAs were conducted on difference scores between twins of a pair to assess whether twins taught together were on average more similar in achievement and motivation than those taught separately. These analyses were conducted on nine measures of achievement and motivation at age 12 in Quebec (Canada) and two measures of achievement at age 16 in the UK. Bonferroni multiple testing correction was set of $p \leq .005$ where $p = .05$ was divided by the number of measures (k = 11) across the two samples.

Results

Frequency Of Separation

Most Quebec twins were in different classes between ages 7 and 12, with only 24-39% taught in the same class, while most UK twin pairs (65.9%) were taught together at age 7, but only 28% were in the same class by age 16 (see Tables 7.2.1, 7.2.2, and 7.2.3). In both samples, the proportion of twin pairs taught together was slightly higher for MZ than DZ twins at all ages. Chi-square tests of separation by zygosity showed no differences in the Quebec sample across all ages. In the UK sample, differences were not present at ages 7, 9, 10, and 14, but at age 12 and 16 more DZ twins than MZ twins were in different classes (age 12: $\chi^2 = 11.967$, $p < .001$; age 16: English, $\chi^2 = 82.564$, maths, $\chi^2 = 51.637$, science, $\chi^2 = 32.854$; $p < .001$). All effect sizes were small, with the greatest effect of 4.4%.

Table 7.2.1. Quebec twin pairs taught by the same or different (S/D) teachers by sex and zygosity; and by zygosity at ages 7 to 12 years

Age	S/D teacher	MZm	DZm	MZf	DZf	DZos	MZ	DZ	Total
Age 7	Different	74.7% n=65	79.1% n=53	70.5% n=67	74.6% n=44	78.8% n=93	72.5% n=132	77.9% n=190	75.6% n=322
	Same	25.3% n=22	20.9% n=14	29.5% n=28	25.4% n=15	21.2% n=25	24.5% n=50	22.1% n=54	24.4% n=104
	Total	100% n=87	100% n=67	100% n=95	100% n=59	100% n=118	100% n=182	100% n=244	100% N=426
Age 9	Different	71.6% n=59	72.9% n=43	63.5% n=54	72.1% n=44	72.4% n=76	77.2% n=125	76.3% n=167	70.3% n=275
	Same	28.4% n=23	27.1% n=16	36.5% n=31	27.9% n=17	27.6% n=29	22.8% n=37	23.7% n=52	29.7% n=116
	Total	100% n=81	100% n=59	100% n=85	100% n=61	100% n=105	100% n=162	100% n=219	100% N=391
Age 10	Different	71.6% n=58	72.9% n=43	63.5% n=54	72.1% n=44	72.4% n=76	67.5% n=122	72.4% n=167	70.3% n=275
	Same	28.4% n=23	27.1% n=16	36.5% n=31	27.9% n=17	27.6% n=29	32.5% n=54	27.6% n=62	29.7% n=116
	Total	100% n=81	100% n=59	100% n=85	100% n=61	100% n=105	100% n=166	100% n=255	100% N=391
Age 12	Different	57.6% n=38	57.4% n=35	59.1% n=52	64.9% n=37	62.2% n=69	58.4% n=90	61.4% n=140	60.3% n=231
	Same	42.4% n=28	42.6% n=26	40.9% n=36	35.1% n=20	37.8% n=42	41.6% n=64	38.6% n=88	39.7% n=152
	Total	100% n=66	100% n=61	100% n=88	100% n=57	100% n=111	100% n=154	100% n=228	100% N=383

MZm = monozygotic male; MZf = monozygotic female; DZm = dizygotic male; DZf = dizygotic female; DZos = dizygotic opposite sex; MZ = all monozygotic; DZ all dizygotic. Significant results in bold at $p \leq .05$.

Table 7.2.2. UK twin pairs taught by the same or different (S/D) teachers by sex and zygosity; and by zygosity at ages 7 to 12 years

Age	S/D teacher	MZm	DZm	MZf	DZf	DZos	MZ	DZ	Total
Age 7	Different	36.5% n=404	37.3% n=393	32.7% n=420	31.7% n=361	33.4% n=702	34.4% n=824	33.9% n=1456	34.1% n=2280
	Same	63.5% n=702	62.7% n=662	67.3% n=866	68.3% n=778	66.6% n=1398	65.6% n=1568	66.1% n=2838	65.9% n=4406
	Total	100% n=1106	100% n=1055	100% n=1286	100% n=1139	100% n=2100	100% n=2392	100% n=4294	100% N=6686
Age 9	Different	42.0% n=238	42.8% n=229	39.4% n=273	40.1% n=234	42.4% n=452	40.6% n=511	41.9% n=915	41.4% n=1426
	Same	58.0% n=328	57.2% n=306	60.6% n=420	59.9% n=350	57.6% n=613	59.4% n=748	58.1% n=1269	58.6% n=2017
	Total	100% n=566	100% n=535	100% n=693	100% n=584	100% n=1065	100% n=1259	100% n=2184	100% N=3443
Age 10	Different	45.6% n=241	49.3% n=252	43.2% n=293	47.1% n=269	46.8% n=504	44.2% n=534	47.5% n=1025	46.3% n=1559
	Same	54.4% n=288	50.7% n=259	56.8% n=386	52.9% n=302	53.2% n=574	55.8% 674	52.5% 1135	53.7% n=1809
	Total	100% n=529	100% n=511	100% n=679	100% n=571	100% n=1078	100% n=1208	100% n=2160	100% N=3368
Age 12	Different	66.4% n=725	67.7% n=710	61.6% n=792	61.5% n=715	71.6% n=1535	**63.8%** **n=1517**	**68.0%** **n=2960**	66.5% n=4477
	Same	33.6% n=367	32.3% n=339	38.4% n=493	38.5% n=447	28.4% n=608	**36.2%** **n=860**	**32.0%** **n=1394**	33.5% n=2254
	Total	100% n=1092	100% n=1049	100% n=1285	100% n=1162	100% n=2143	**100%** **n=2377**	**100%** **n=4354**	100% N=6731

MZm = monozygotic male; MZf = monozygotic female; DZm = dizygotic male; DZf = dizygotic female; DZos = dizygotic opposite sex; MZ = all monozygotic; DZ all dizygotic. Significant results in bold at $p \leq .05$

Table 7.2.3. UK twin pairs taught by the same or different (S/D) teachers by sex and zygosity; and by zygosity at ages 14 to 16 years

Age	S/D teacher	MZm	DZm	MZf	DZf	DZos	MZ	DZ	Total
Age 14	Different	79.4% n=108	73.9% n=88	70.3% n=130	71.8% n=112	78.9% n=195	74.1% n=238	75.7% n=395	75.1% n=633
	Same	20.6% n=28	26.1% n=31	29.7% n=55	28.2% n=44	21.1% n=52	25.9% n=83	24.3% n=127	24.9% n=210
	Total	100% n=136	100% n=119	100% n=185	100% n=156	100% n=247	100% n=321	100% n=522	100% N=843
Age 16 English	Different	67.8% n=202	80.7% n=230	63.5% n=216	72.5% n=240	84.8% n=530	**65.5% n=418**	**80.6% n=1000**	75.5% n=1418
	Same	32.2% n=96	19.3% n=55	36.5% n=124	27.5% n=91	15.2% n=95	**34.5% n=220**	**19.4% n=241**	24.5% n=461
	Total	100% n=298	100% n=285	100% n=340	100% n=331	100% n=625	**100% n=638**	**100% n=1241**	100% N=1879
Age 16 Maths	Different	63.5% n=190	76.3% n=219	55.3% n=188	76.5% n=254	81.5% n=507	**59.2% n=378**	**79.0% n=980**	72.2% n=1358
	Same	36.5% n=109	23.7% n=68	44.7% n=152	23.5% n=78	18.5% n=115	**40.8% n=261**	**21.0% n=261**	27.8% n=522
	Total	100% n=299	100% n=287	100% n=340	100% n=332	100% n=622	**100% n=639**	**100% n=1241**	100% N=1880
Age 16 Science	Different	65.7% n=195	72.7% n=208	53.2% n=181	63.7% n=211	76.2% n=475	**59.0% n=376**	**72.1% n=894**	67.7% n=1270
	Same	34.3% n=102	27.3% n=78	46.8% n=159	36.3% n=120	23.8% n=148	**41.0% n=261**	**27.9% n=346**	32.3% n=607
	Total	100% n=297	100% n=286	100% n=340	100% n=331	100% n=623	**100% n=637**	**100% n=1240**	100% N=1877

MZm = monozygotic male; MZf = monozygotic female; DZm = dizygotic male; DZf = dizygotic female; DZos = dizygotic opposite sex; MZ = all monozygotic; DZ = all dizygotic. Significant results in bold at $p \leq .05$.

Average Effects Of Classroom Separation

Means and standard deviations (SD) for all assessed variables at ages 7, 9, 10, 12, 14, and 16 by same or different teacher can be found in Appendix 6, Tables 6.2 to 6.20 for the whole sample, the five sex by zygosity groups (MZm, DZm, MZf, DZf, DZos), and zygosity (MZ, DZ).

The patterns of results were very similar for twins taught separately and together across zygosity groups. ANOVAs (presented in Tables 7.3.1, 7.3.2 and 7.3.3 for achievement, and Table 7.4 for motivation) showed no differences for most measures in achievement, cognitive ability, motivation and teacher-student relations between same vs. different class groups. A few differences were found, although with very weak effect sizes (ranging from 0.2% to 2.8%). The biggest effect of 2.8% was observed for maths GCSE (UK), with twins taught in the same class performing better than those in different classes. Levene's tests revealed unequal variance for these analyses, however, with the smallest amount of variance revealed for MZ twins in a different classroom (0.72) vs. MZ twins in the same classroom (0.96). DZ twins in the same classroom had the smaller amount of variance (0.77), vs. DZ twins in different classrooms (0.90). Results of ANOVA for sex and zygosity are presented in Tables 7.3.1, 7.3.2 and 7.3.3 and 7.4. Although some significant differences between the sex by zygosity groups (MZm, DZm, MZf, DZf, DZos) were found, after correction for multiple testing only a few differences were present, all of weak effect size (ranging from 0.3% to 3.1%). Levene's tests indicated equal variances were assumed for the majority of analyses. Where unequal variances were revealed, they are indicated in Tables 7.3 to 7.4.

Table 7.3.1. Achievement: ANOVA results at ages 7 and 9 by zygosity, sex and by having the same or different (S/D) teachers

Age	Country	Construct	School subject	S/D teacher		Zygosity*S/D		Sex, Zygosity*S/D		Zygosity*Sex	
				p	η^2	*p*	η^2	*p*	η^2	*p*	η^2
Age 7	Quebec-Canada	Achievement	Reading	.579	.001	.103	.008	.217	.030	.127	.013
			Writing	.197	.005	.154	.006	.448	.022	.221	.010
			Maths	.292	.004	.439	.002	.574	.018	.539	.004
			In general	.273	.004	.123	.008	.173	.033	.234	.009
		T/S relation		.496	.001	.833	.000	.050^	.015^	.690^	.002^
	UK	Achievement	Maths	.098^	.001^	.737^	.000^	.364^	.001^	.003^	.003^
			English	.053	.001	.965	.000	.059	.002	**.000**	**.012**
			Non-verbal	.224	.000	.287	.000	.781	.000	.004	.003
			Verbal	.124	.001	.079	.001	.184	.001	.002	.003
			g	.897	.000	.091	.001	.322	.001	**.000**	**.005**
			Reading	.319	.000	.100	.001	.300	.001	**.000**	**.005**
Age 9	Quebec-Canada	Achievement	Reading	.797	.000	.617	.001	.309	.022	.349	.006
			Writing	.317	.003	.448	.002	.170	.028	.232	.008
			Maths	.340	.002	.246	.004	.248	.025	.181	.009
			Sciences	.740	.000	.580	.001	.605	.015	.231	.008
			In general	.650	.001	.394	.002	.321	.022	.347	.006
		T/S relation		.803	.000	.085	.008	.968	.000	.114	.012
	UK	Achievement	Maths	.179	.001	.895	.000	.929	.000	.050	.004
			English	.061	.001	.461	.000	.732	.001	**.000**	**.013**
			Non-verbal	.170	.001	.276	.000	.058	.003	.138	.002
			Verbal	.009	.002	.371	.000	.585	.001	.050	.003
			g	.018	.012	.874	.000	.171	.002	.026^	.004^

Bold indicates significance with a Bonferroni multiple testing correction applied of $p = .05$ divided by the number of measures (k=76) across all ages (7 to 16) across both samples, providing a significance value of $p \leq .001$ (.05/76). T/S relation = teacher-student relationship. ^ = unequal variance: Levene's test significant at $p \leq .05$.

Table 7.3.2. Achievement: ANOVA results at ages 10 and 12 by zygosity, sex and by having the same or different (S/D) teachers

Age	Country	Construct	School subject	S/D teacher		Zygosity*S/D		Sex, Zygosity*S/D		Zygosity*Sex	
				p	η^2	*p*	η^2	*p*	η^2	*p*	η^2
Age 10	Quebec-Canada	Achievement	Reading	.497	.001	.107	.009	.425	.018	.189	.009
			Writing	.315	.003	.642	.001	.027	.041	.009	.025
			Maths	.777	.000	.964	.000	.221	.025	.073	.014
			In general	.678	.000	.315	.003	.194	.026	.023	.020
		T/S relation		.187^	.004^	.333	.002	.424	.005	.582	.003
	UK	Achievement	Maths	.106	.001	.580	.000	.472	.001	.052	.003
			English	.585^	.000^	.787^	.000^	.370	.002	**.000**	**.009**
			Non-verbal	.261	.001	.901	.000	.886	.001	.018	.005
			Verbal	.601^	.000^	.547	.000	.546^	.001^	**.000^**	**.016^**
			g	.301	.000	.890	.000	.800	.001	**.000**	**.013**
			Reading	.454^	.000^	.379	.000	.182^	.002^	.105^	.003^
Age 12	Quebec-Canada	Achievement	Reading	.026	.016	.063	.011	.174^	.034^	.722	.002
			Writing	.014	.019	.490	.002	.268	.029	.290	.008
			Maths	.004	.027	.097	.009	.481	.022	.272^	.009^
			In general	.003	.029	.093	.009	.216	.032	.117	.014
		T/S relation		.417	.002	.916^	.000^	.042^	.021^	.880^	.001^
	UK	Achievement	Maths	.672	.001	.113	.001	.073	.002	.442	.001
			English	.076^	.001^	.717^	.000^	.357^	.001^	**.000^**	**.009^**
			Non-verbal	.289	.000	.296	.000	.258	.001	.084	.002
			Verbal	.040	.001	.888	.000	.194	.001	**.000**	**.014**
			g	.033	.001	.482	.000	.094	.002	**.000**	**.009**
			Reading	.917	.000	.955	.000	.993^	.000^	.121	.002

Bold indicates significance with a Bonferroni multiple testing correction applied of $p = .05$ divided by the number of measures (k=76) across all ages (7 to 16) across both samples, providing a significance value of $p \leq .001$ (.05/76). T/S relation = teacher-student relationship. ^ = unequal variance: Levene's test significant at $p \leq .05$.

Table 7.3.3. Achievement: ANOVA results for the UK twins at ages 14 and 16 by zygosity, sex and by having the same or different (S/D) teachers

			S/D teacher		Zygosity*S/D		Sex, Zygosity*S/D		Zygosity*Sex	
Age	**Construct**	**School subject**	***p***	**η^2**	***p***	**η^2**	***p***	**η^2**	***p***	**η^2**
Age 14	Achievement	Maths	.030^	.010^	.967^	.000^	.904^	.002^	.651^	.005^
		English	.025^	.011^	.941^	.000^	.979	.001	.236	.012
		English reading	.577	.002	.394	.004	.627	.013	.876	.006
		Non-verbal	.706^	.000^	.577	.000	.013^	.005^	.595^	.001^
		Verbal	.123	.001	.520	.000	.908	.000	.217	.002
		g	.355	.000	.547	.000	.272	.002	.413	.002
Age 16	Achievement	Maths GCSE	**.000^**	**.028^**	.104^	.002^	.469	.002	.207	.001
		English GCSE	**.000**	**.008**	.303	.001	.180	.004	**.000**	**.019**
		English lang GCSE	**.000**	**.009**	.172	.001	.249	.003	**.000**	**.016**
		English lit GCSE	.081	.002	.774	.000	.559	.002	**.000**	**.016**

Bold indicates significance with a Bonferroni multiple testing correction applied of p = .05 divided by the number of measures (k=76) across all ages (7 to 16) across both samples, providing a significance value of $p \leq .001$ (.05/76). lang = language; lit = literature. ^ = unequal variance: Levene's test significant at $p \leq .05$.

Table 7.4. Motivation: ANOVA results from ages 9 to 12 by zygosity, sex and being taught by the same or different (S/D) teachers

Age	Country	Construct	School subject	S/D teacher *p*	S/D teacher η²	Zygosity*S/D *p*	Zygosity*S/D η²	Sex, Zygosity*S/D *p*	Sex, Zygosity*S/D η²	Zygosity*Sex *p*	Zygosity*Sex η²
Age 9	UK	Enjoyment	English	.277	.000	.344	.000	.608	.001	**.000**	**.001**
			Maths	.444	.000	.128	.001	.421	.001	**.000**	**.019**
		SPA	English	.008	.008	.832	.000	.091	.003	**.000**	**.007**
			Maths	.046	.001	.765	.000	.826	.001	**.000**	**.021**
		Motivation	Academic O	.018^	.002^	.319^	.000^	.751	.001	**.000**	**.008**
Age 10	Quebec-Canada	Enjoyment	Reading	.976	.000	.494	.001	.053	.036	.843	.001
			Maths	.445^	.002^	.406^	.002^	.922^	.007^	.896^	.001^
		SPA	Reading	.202^	.005^	.757	.000	.346^	.022^	.929	.000
			Maths	.850	.000	.192	.005	.753	.012	.622	.003
Age 12	Quebec-Canada	Enjoyment	Reading	.403	.002	.659^	.001^	.010	.026	.024	.021
			Maths	.032	.013	.584	.001	.563	.003	.338	.006
		SPA	Reading	.291	.003	.151	.006	.176	.010	.288	.007
			Maths	.012	.018	.418	.002	.555	.003	.924	.000
	UK	Enjoyment	English	.006^	.002^	.808	.000	.054^	.002^	**.000^**	**.032^**
			Maths	.004	.002	.362	.000	.839^	.000^	.101	.002
			Academic	**.001**	**.002**	.903	.000	.063	.002	.068	.002
		SPA	English	.037^	.001^	.076^	.001^	.303^	.001^	**.000^**	**.008^**
			Maths	.006	.002	.901	.000	.992	.000	**.000**	**.014**
			Academic	.003	.002	.224	.000	.537	.001	.516	.001
		Motivation	Academic O	**.000**	**.003**	.567	.000	.276	.001	.896	.000

Bold indicates significance with a Bonferroni multiple testing correction applied of p = .05 divided by the number of measures (k=76) across all ages (7 to 16) across both sample providing a significance value of $p \leq .001$ (.05/76). Academic O = Academic overall. ^ = unequal variance: Levene's test significant at $p \leq .05$.

Within-Pair Similarity Of Twins Taught Together Or Separately

Because some weak effects of separation were suggested at age 12 (Quebec) and 16 (UK), additional ANOVAs were conducted at these ages to test whether twin pairs taught together were more similar to each other than those taught separately. The difference in scores between twin and co-twin in each pair were computed for all constructs of the Quebec sample at age 12; and for maths and English GCSE grades of the UK sample at age 16. Using the within-pair difference scores, ANOVAs were conducted by same vs. different classrooms and zygosity; and by same vs. different classrooms and sex by zygosity (see Table 7.5).

Overall, the results showed smaller mean difference scores for the twins taught together than separately (see Appendix 6, Tables 6.21 to 6.23). In other words, within-pair similarity was greater for twin pairs taught together than apart. Figures 7.1, 7.2, 7.3 and 7.4 show within-pair differences (or similarity) by zygosity and same vs. different classrooms. Greater within-pair similarity was found for MZs than DZs, with the greatest within-pair difference shown for DZs taught separately. However, only a few of the differences reached significance after correction for multiple testing: English and maths GCSE at age 16 (UK), with larger differences seen for separated DZ twins but with small effects (2.2% to 4.2%). Small significant differences were found between sex and zygosity groups, after correction for multiple testing, but these did not differ as a function of same different classroom (see Table 7.5). Levene's tests indicated equal variances were assumed for the majority of analyses in the Quebec sample. However, unequal variances were revealed for the UK analyses, and are indicated in Table 7.5. For English GCSE, the smallest amounts of variance were revealed for twins in the same classroom and the largest for twins in different classrooms: MZ twins, same classroom (0.34), vs. MZ twins, different classroom (0.49); and DZ twins, same classroom (0.58) vs. DZ twins, different classroom (0.92). For maths GCSE a similar pattern was observed, with the smallest amounts of variance revealed for twins in the same classroom and the largest for twins in different classrooms: MZ twins, same classroom (0.29), vs. MZ twins, different classroom (0.50); and DZ twins, same classroom (0.46) vs. DZ twins, different classroom (0.83).

Table 7.5. ANOVA for difference scores between twin pairs taught by the same or different teachers by zygosity, sex and being taught by the same or different (S/D) teachers ages 12 and 16

Age	Country	Construct	School subject	S/D teacher		Zygosity * S/D		Sex, Zygosity*S/D		Zygosity*Sex	
				p	η^2	*p*	η^2	*p*	η^2	*p*	η^2
Age 12	Quebec-Canada	Achievement	Reading	.010	.026	.421^	.003^	.874	.005	.006	.055
			Writing	.013	.024	.042^	.016^	.159	.026	.019	.046
			Math	.243	.006	.975^	.000^	.209	.023	**.000**	**.116**
			In general	.085	.012	.928	.000	.367	.017	**.000**	**.080**
		T/S relation		.009^	.019^	.514^	.001^	.841	.004	.052	.026
		Enjoyment	Reading	.210	.005	.569^	.001^	.713	.006	**.000**	**.033**
			Math	.222	.004	.462^	.002^	.721	.006	.009	.039
		SPA	Reading	.180	.005	.148	.006	.559	.009	.462	.011
			Math	.384	.002	.065^	.010^	.254	.016	**.002**	**.048**
Age 16	UK	Achievement	Maths GCSE	**.001^**	**.042^**	.008^	.004^	.099^	.005^	**.000^**	**.039^**
			English GCSE	**.000^**	**.022^**	.338^	.001^	.734^	.001^	**.000^**	**.033^**

Bold indicates significance with a Bonferroni multiple testing correction applied of *p* = .05 divided by the number of measures (k=11) across ages 12 and 16 and across both samples which provided a significance value of *p* ≤ .005 (.05/11).T/S relation = teacher-student relationship. ^ = unequal variance: Levene's test significant at *p* ≤ .05.

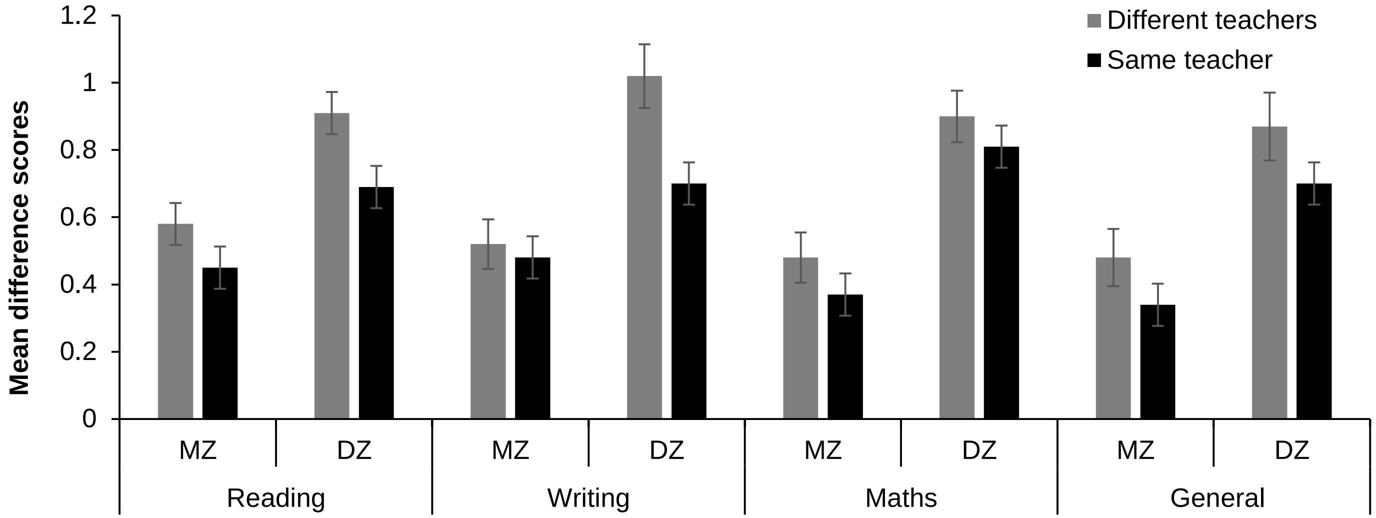

Figure 7.1. Raw mean difference scores in reading, writing, maths and general achievement at age 12 for Quebec MZ and DZ twin pairs taught by the same or different teachers

Not significant after correction for multiple testing $p \leq .005$ (.05/11)

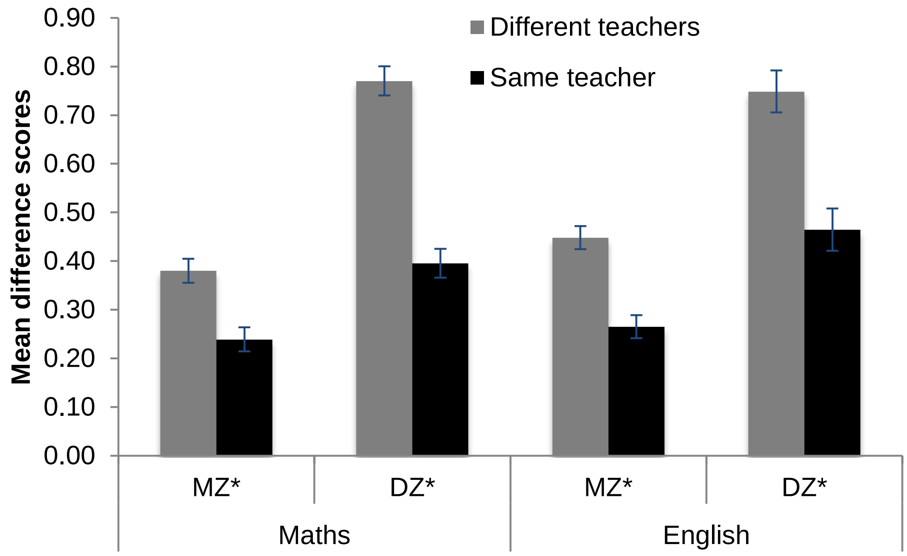

Figure 7.2. Raw mean difference scores for GCSE grades in maths and English at age 16 for UK MZ and DZ twin pairs taught by the same or different teachers

* = Significant differences found following correction for multiple testing $p \leq .005$ (.05/11)

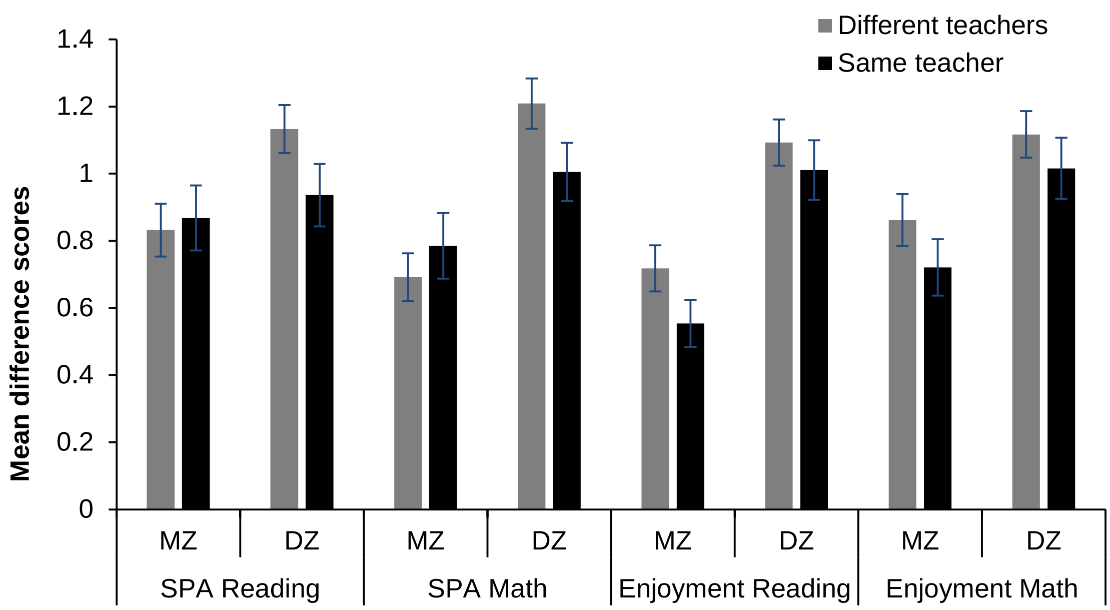

Figure 7.3. Raw mean difference scores for self-perceived ability and enjoyment of reading and maths at age 12 for Quebec MZ and DZ twin pairs taught by the same or different teachers

Not significant after correction for multiple testing $p \leq .005$ (.05/11)

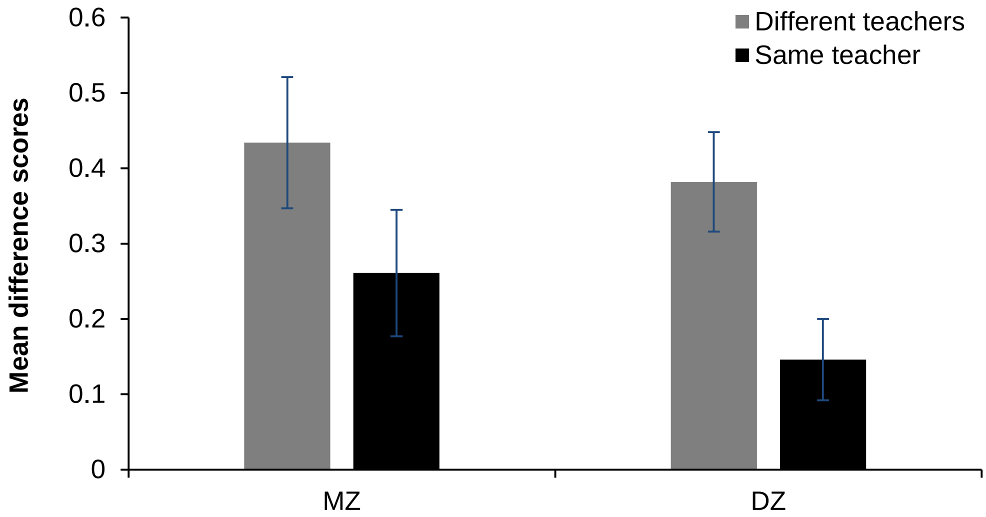

Figure 7.4. Raw mean difference scores for teacher-student relations at age 12 for Quebec MZ and DZ twin pairs taught by the same or different teachers

Not significant after correction for multiple testing $p \leq .005$ (.05/11)

Cumulative Effect Of Separation

Additional analyses were performed to test whether there was an accumulative effect of classroom separation on twins' achievement and motivation across their years of education from age 7. Appendix 6, Tables 6.24 and 6.25 present the percentage of twins who were educated in the same classrooms most of their school years vs. twins in different classes most of their school years. The following analyses were conducted on maths and English at age 16 by twins taught together or separately for most of the time up to age 14. ANOVAs conducted at ages 12 (Quebec) and 16 (UK) revealed no significant differences between these groups after correcting for multiple testing ($p \leq .005$, .05/11, where k = 11).; this was the case for both MZ and DZ twins (see Appendix 6, Table 6.26). Levene's tests indicated equal variances were assumed for the majority of analyses. Where unequal variances were revealed, they are indicated in Appendix 6, Table 6.26.

Cross-Cultural Generalizability Across The Two Education Systems

The results of the present investigation highlight both similarities and differences in classroom separation between the samples from Canada and the UK. The Quebec sample shows a greater proportion of twins taught separately at the beginning of elementary school than at the end (at age 12), while in the UK, a greater proportion of twins are taught together in elementary school than in high school (ages 12 to 16) (see Tables 7.2.1, 7.2.2, and 7.2.3). By age 12, the proportions of twins taught separately are similar across the two countries. Despite some differences in separation practices across school years, the present study revealed no effect of classroom separation on school achievement, cognitive ability and motivation in both Quebec and UK.

Discussion

The main aim of the present study was to examine the effect of classroom separation on school achievement, cognitive ability, motivation of twins and teacher-student relations. The study found almost no differences between twins taught together and those taught separately for any of the measures. These results are consistent across ages and countries as no separation effects were found for ages 7 to 12 in Quebec-Canada and ages 7 to 14 in the UK (see Tables 7.3.1, 7.3.2, 7.3.3 and 7.4). The only differences found at age 16 (UK) showed a weak effect (see Table 7.3.3), in favour of educating twins together. These results are also consistent across sex and zygosity as no effects of separation were found for any specific sex and zygosity groups. The study also found no cumulative effect of separation across years of education (see Table 6.26 in Appendix 6). The Levene’s tests also revealed unequal variances where differences were observed in the UK sample, which may compromise interpretation of these results.

These findings corroborate previous research that found no significant differences between twin pairs taught together or separately for school achievement (Coventry et al., 2009; Kovas et al., 2007; Polderman et al., 2009); cognitive abilities (Kovas et al., 2007); and academic motivation (Kovas et al., 2015). The findings are also consistent with previous research that found no cumulative effect of separation (Kovas et al., 2015; Webbink et al., 2007).

These results also offer some support for a previous study that showed greater within-pair similarity for twins taught together vs. twins taught separately, with greater similarity for MZ twins than DZ twins (Byrne et al., 2010). Indeed, the study found slightly greater within-pair similarity for twins taught together with slightly more within-pair similarity found for MZ twins. These results were only found at age 16 (UK) and with very modest effects (see Table 7.5).

Previous research indicated effects of classroom separation might be stronger for earlier school years compared with later school years (Tully et al., 2004; Webbink et al., 2007). The present study did not replicate this: the absence of classroom separation effects was consistent across ages (see Tables 7.3.1, 7.3.2, 7.3.3 and 7.4).

Overall, although some studies found significant effects of classroom separation, well-powered studies found negligible or small effects of classroom separation. Inconsistencies in previous studies could be due to differences in samples (e.g., spurious effects in unrepresentative samples) (see Table 7.1). Another explanation for the non-significance of classroom separation is the possibility that other aspects of the classroom environment, such as quality of instruction or peer relations, may buffer any effect of separation on achievement (e.g. Hamre & Pianta, 2005). These may also explain our non-significant findings for teacher-student relations between twins taught together and twins taught separately. It may also be that, as twins' classroom allocation is usually a result of discussion between parents, teachers and the twins themselves, any potential ill effects of assignment may be attenuated, and could potentially be present only if decisions were determined by high-level school policy beyond family and teacher control.

This study shows a highly similar pattern of results for achievement and motivation across the two samples for ages 7 to 12 years. This finding is surprising in light of differences between the two samples regarding timing and frequency of classroom separation. In Quebec (Canada) a greater proportion of twins are taught separately at the beginning of elementary school than at the end (age 12). In the UK, the reverse situation occurs: a greater proportion of twins are taught together during their entire elementary education (up to age 11). This likely reflects differences in educational policies for the two countries. In Quebec, the School Commission Boards strongly encourage separation of twins when they begin education (Lalonde & Moisan, 2003) whereas separation in the UK occurs later on in secondary education/high school, potentially as a result of ability selection.

It is possible that previously reported effects of separation resulted from setting and streaming by ability processes rather than any effect of separation per se. Indeed, significant effects were found at age 16 in the UK where students are streamed for ability. In contrast, in Quebec, where separation effects were negligible, there is no ability streaming. UK twin pairs at this later stage of their education are more likely to be taught separately as a result of different subject choices and differences in ability. This is particularly true of DZ twins as they are usually less similar phenotypically than MZ twins (Petrill et al., 2009; Spinath, Spinath et al., 2006) and therefore end up in separate classrooms more often than MZ twins. This study did indeed find larger numbers of DZ than MZ twin pairs taught separately at age 16 in the UK, whereas the numbers were similar across zygosity groups for prior years in Quebec and the UK (see Tables 7.2.1, 7.2.2 and 7.2.3). The difference in classroom separation between DZ and MZ twins was slightly larger for maths (DZ 79% vs. MZ 59.2%) than English (DZ 80.6% vs. MZ 65.5%). The present study also found a marginally greater effect of separation for maths (2.8%) than for English (0.8%) (see Table 7.3.3); and a slightly larger effect of separation on within-pair similarity for maths (4.8%) than for English (2.2%) (see Table 7.5). These differences are small and suggest a trend that may be explained by the greater genetic overlap found for intelligence with maths GCSE than with other GCSE subjects (Rimfeld, Kovas, Dale & Plomin, 2015).

Although the study did not find major differences as a result of classroom separation, this does not mean an absence of effect for the individual. Effects of classroom separation are likely to depend on individual characteristics and different perceptions of classroom experience (e.g. Asbury, Almeida, Hibel, Harlaar, & Plomin, 2008). This is clearly demonstrated by Figures 6.2 and 6.3

in Appendix 6 that show difference scores in motivation for individual twin pairs taught together at ages 9, 10 and 12 compared with twin pairs taught together at age 9 and 10, then separately at age 12. If we anticipate a strong effect of separation we should expect to see few positive scores, indexing similarity between twins, and a larger number of negative values, indicating differences between twins, for twin pairs with different teachers. Equally, we would expect a larger number of positive values, and few negative or unchanged scores for twin pairs with the same teacher. Instead, scores for both groups of twin pairs are normally distributed with many individuals with positive and negative scores. For some, enjoyment/perceived ability increases, for others it goes down. This suggests that individual differences play a large role in perceptions of classroom experience. Indeed, separation may not even be a factor; there may be other influences, for example classroom/teacher effects, ability streaming or a change of school that occurs at age 11 for UK students.

Limitations And Future Research

The study is not without limitations and one drawback is unavailable data at age 16 in the Quebec sample to fully test the patterns across both samples. Another limitation is that at age 16 in the UK sample, data were only available for GCSE results in the sub-sample that provided same/different teacher data. Therefore the study was unable to assess any effect of same/different teacher on cognitive ability or motivation at this age. One other weakness is attrition for both samples that resulted in some non-overlapping data across the years of the study and so prevented further longitudinal analyses to show potential causal effects for consecutive years of being in the same vs. different classes. It is also worth bearing in mind that although there are some differences between the two education systems, the cultures investigated here are very similar. Future studies would benefit from investigating across more diverse cultures and education systems. Finally, the teachers reported twins' school achievement. It is possible that teachers rated twins in the same classroom more similarly. However, the significant effect of classroom separation on achievement at age 16 (UK) is unlikely to be the result of inter-rater reliability; as at this age, achievement was measured by externally assessed exams. If differences were due to teacher rating, it would more likely occur in the younger ages where the twins' teachers assessed the twins' ability, and especially in the UK where the class teachers set and marked the national curriculum tests on which these assessments were based.

Conclusion

These results show no sizeable positive or negative average effect of separation on twins' achievement, cognitive ability and motivation. The few effects found were weak and likely to be explained by education selection processes, such as ability streaming, rather than any real effect of classroom separation. This is borne out by the timing of significant effects at age 16 (UK), slightly less similarity found for separated DZ than MZ twin pairs at age 16 (UK), and larger numbers of separated DZ than MZ twin pairs. These results suggest that in terms of academic achievement, cognitive ability and motivation, policymakers should not impose rigid guidelines for schools and parents to separate twin pairs during their education. The choice of whether to educate twin pairs together or separately should be up to parents, twins and teachers, in response to twins' individual needs.

Chapter 8

General discussion, implications and future directions

The present thesis set out to investigate teacher/classroom effects on several educational outcomes. The thesis explored whether students' motivation, academic performance and perception of learning environment were affected by their teachers and/or classmates, as reflected in average differences between classes. Many people believe that teachers differ greatly in their ability to motivate and educate children, and that therefore differences between classes are due to teacher differences. This assumed 'effect of the teacher' implies that an individual child may be performing below their potential BECAUSE of the teacher's shortcomings. Equally, many people attribute their interest or success in a particular subject to excellence of their teacher, implying that they would be less interested or worse performing were they taught by a different teacher. However, these assumptions and beliefs remain largely untested as true experiments in education are rarely possible – researchers do not have *full control* over independent variables or *truly random* allocation to experimental conditions. Therefore, it remains unclear how much 'value is added' by teachers and classrooms on top of an individual student's characteristics and circumstances.

This thesis reviewed the available evidence on teacher and classroom effects and reports further original research, employing unique pseudo-experimental methods. The results led to 3 main conclusions regarding teacher/class effects: (1) the effects, when present, are weak; (2) the strength of effects depends on educational settings; (3) the effects of the current teacher are confounded by effects of previous teachers, classroom composition, student ability and above all – formal or hidden selection practices. These conclusions have direct implications for education: praising or blaming teachers for average performance of the classes they teach, which is flawed practice. Fixation on evaluation of average performance distracts from true challenges facing education, namely continuous scientific pursuits in search of new educational tools to help teachers in their work.

In the original research reported in this thesis classroom and teacher groups were compared across two domains, maths and geography, for students in their first year of secondary education (aged 10 to 12 years) with specific subject teachers for the first time. Using a longitudinal design, educational outcomes were explored to see how they unraveled across several assessments during one academic year. A cross-cultural approach was used, which allowed the comparison between two different education systems, in Russia and the UK (Chapters 3, 4, 5 and 6). In so doing, the research was able to take account of the differences in streaming and tracking processes between the two education systems. In the UK sample, streaming was applied for maths classes, whereas, in the Russian sample, no formal selection processes took place. Assessing several factors simultaneously allowed a more fine-grained approach to investigate teacher/classroom effects than approaches used previously in the literature.

Another pseudo-experimental study, conducted as part of this thesis, compared motivation and achievement of twins taught together and twins taught separately (Chapter 7). The data came from two large twin samples from the UK and Quebec (Canada), whose participants were followed longitudinally from ages 7 to 16 years and assessed on measures of school achievement, motivation and cognitive ability. Differences in education systems were considered, as unlike the UK, the education system in Quebec does not formally apply streaming and tracking. Greater differences between twins if taught separately over those taught together could imply an effect of teacher/classroom.

The first chapter provided an overview of key research in the literature that investigated teacher/classroom effects. It introduced the different approaches applied previously in the field, such as random allocation to a class and teacher group, econometric studies that used large-scale survey data, and behavioural genetic research. It also considered different factors suggested to contribute towards student achievement, such as class size, classroom composition/streaming, and teacher characteristics. The chapter identified several research avenues regarding differences between class and teacher groups in an attempt to disentangle contributing factors that are often erroneously assumed to be ‘teacher/class effects’. In order to provide evidence to address these questions, the empirical chapters 3, 4, 5, 6, and 7 consider similarities and differences in educational outcomes across countries/education systems; classroom and teacher groups; and twin pairs taught together or separately; as well as associations between teacher characteristics and educational outcomes.

As many of the measures analysed in these chapters were originally devised for different ages to the present sample of 10 to 12 year olds, some adaptation in line with their curriculum was needed. Additionally, measures were originally developed to assess mathematics so further modification was required to assess geography. The measures were adapted and developed during the pilot study, reported in Chapter 2. The pilot study, conducted in the UK, also tested the timing and validity of the revised measures for students and feasibility of measures for teachers. The conclusions from the pilot study informed further adaptations of the measures as well as the data collection procedure.

Chapter 3 addressed whether there were differences between the two countries in academic outcomes, given a number of differences in education systems. For example, streaming by ability was employed in the UK sample's maths classes, whereas no formal selection took place in Russia. However, the Russian sample attended schools with enhanced language curricula that provided the opportunity to learn up to two second languages: English; English and Spanish, or English and Chinese. The curriculum of one school in the UK sample had a music and ICT (information technology) focus and the other school had a broad general curriculum. Students in Russia had a much longer summer break compared to the UK students, although both samples completed the same number of days in school across the academic year. The results showed that the two samples were similar across the academic year for almost all maths and geography educational outcomes. They were also similar in perceptions of intelligence and socioeconomic status. Both samples are from mixed ability schools, but streaming occurs in the UK for maths classes. It may be that informal selection processes in the Russian sample increased their similarity further. As students had the opportunity to learn one or two languages, parents who had elected to enroll their child into the more challenging linguistic curriculum (learning English and Spanish or English and Chinese), would likely be more confident in their child's ability, perhaps more so than parents who had chosen one language programme. This may mean some stratification within the schools regarding ability, similar to that occurring in the UK maths classes, and therefore the samples were more closely matched. The resemblance between the two samples provided a good basis from which to make further within group comparisons. The samples also appeared to be largely representative of their countries, an important factor when making inferences regarding different education systems. The study concluded that the two education systems lead to similar educational outcomes, and that factors that drive individual differences within populations are likely to be similar in the UK and Russia.

Chapter 4 focused on the Russian sample which provided some unique opportunities for examination of potential effects. The children remained in the same school throughout their education, within the same class groups. They also had the same primary school teacher for the entire four years of primary education from age 7 to 11 years. The study investigated whether being in the same class with the same peers during primary and secondary education would lead to a significant effect of teacher/classroom on measures of school achievement, performance, classroom environment, motivation and subject anxiety at time 1. The study also explored whether influences from the primary school classroom had an overriding effect beyond that of the current subject teacher.

For the majority of measures assessed at time 1 no significant effect of teacher/classroom was found. However, being among the same peers for the previous four years and remaining within the same class group had a moderate effect on school achievement, classroom environment, student-teacher relations and classroom chaos. A comparison between current subject teacher groups, whereby students are regrouped according to their maths or geography teachers, showed similar but more modest effects for the same measures. This reduction in effects for current subject teacher group, suggests that other factors contribute beyond the current subject teacher.

In order to establish any influence from the primary school teacher beyond the influence of the current subject teacher, associations were estimated between corresponding measures across the two domains, for example, between maths performance and geography performance or between maths classroom chaos and geography classroom chaos. Strong correlations across subjects (maths and geography) that were taught by different teachers, would be consistent with the influence of primary school teacher; weak correlations would be consistent with a stronger influence of current subject teacher. The results showed moderate to strong correlations, indicating a negligible influence of current subject teacher. They were not so strong however, to signify an overriding influence of primary school teacher. The findings also suggest that being among the same peers with the same primary school teacher for four years of education has some influence beyond the modest effect of the current subject teacher. The correlations across subjects, however, may simply reflect ‘g’, or indeed other general genetic, neural, cognitive, or motivational factors: mechanisms that are unrelated to classes or teachers. For example, the strong correlations (around .70) between reading, maths, and science (Krapohl et al., 2014), have been shown to be largely due to substantial genetic overlap (e.g. .74 between reading and maths) across different domains (Kovas, Harlaar, Petrill & Plomin, 2005), in line with the ‘generalist genes’ hypothesis (Kovas & Plomin, 2006).

If the primary school teacher was setting a class ethos that was unchangeable by the current subject teacher it would be expected to see consistent patterns of class ranking across the measures that showed a significant effect of class or current subject teacher. The level of correspondence in rank across the measures implies some influence of primary school teacher. However, as complete consistency in rankings was not found across measures and domains for all class and teacher groups, it suggests other factors are involved. For example, implicit selection processes may be playing a role, with higher ability students being drawn towards learning two languages instead of one. The study tested this possibility by comparing one language vs. two language classes in the same school in Russia. However, expected differences between the one-language group and the two-language groups were only shown for the English and Spanish linguistic group, and not with the English and Chinese group, suggesting that learning two languages per se, is not driving the differences. Instead, it was established that some high ability students in the English and Spanish group might be driving the effects for one class and teacher group that were consistently observed in the higher ranks. Another potential reason for implicit selection may be a result of parental preferences for specific teachers. Parents may manipulate the class/teacher allocation process by endeavoring to obtain a popular teacher for their child, especially if the child has an older sibling at the school. If such manipulation is performed by particularly motivated/active parents, their children may on average show more motivation or higher ability. The effects shown for specific class/teacher groups and unequal variances for specific groups also indicate the influence of variation in ability. Overall, the study suggests a weak effect of current subject teacher, confounded by multiple factors, many of which stem from primary school.

In chapter 5, the investigation from chapter 4 was taken further to establish continuity of effects of maths and geography classroom teachers across the academic year. Specifically, whether effects found at time 1(the first assessment wave in January) for achievement, performance, classroom environment, student-teacher relations and classroom chaos persisted across the academic year at time 2 (April/May) and time 3 (September, following the summer break). The investigation also took account of primary school achievement to control for primary school influences.

The expected large drop in effect sizes for classroom and increase in effects for teacher groups was observed only for maths performance at time 1, when controlling for primary school achievement. At time 2, after controlling for primary school achievement, no consistent change in effects was observed apart from significant effects for maths and geography year 5 school achievement. Together these findings indicate a stronger influence of primary school at time 1 that weakens by time 2. However, there may be other potential reasons for the change in effects. It is often assumed that increases and decreases in performance during the transition to secondary school are mostly due to change in teachers and schools. However, children at this age undergo many physiological and psychological changes. Moreover, as material gets more complex, children rely on different neural and cognitive resources, which in turn are driven by genetic and environmental factors. New genes come ‘on-line’ to support learning. Furthermore, many changes can happen in children’s lives (parents divorcing, illnesses, improvement in financial circumstances of families) etc., all of these may lead to changes and are confounded in ‘teacher’ effects. However, the measures showing significant effects are related to classroom environment, therefore, it may indeed be an influence of primary school ethos impacting on the classroom environment of the current subject teacher. Overall, the findings suggest a weak influence from primary school that extended across the following academic year into secondary education and contributed towards variation in student outcomes between classroom and teacher groups. Although weak, the influence remained more pronounced for geography classrooms than for maths.

In comparing results of the Russian sample with the UK sample where formal selection processes were employed for maths but not geography, the Russian sample results appear to fit somewhere between the results of the maths and geography classrooms for the UK. The degree of consistency in ranking patterns for the Russian sample is far greater than those observed in the UK sample geography classrooms - which should be the most comparable. The ranking patterns are less consistent however, than those shown for the UK sample maths classrooms. The number of significant effects is largely equal across samples, although effect sizes differ considerably.

As expected with classes streamed by ability, achievement and performance in the UK sample showed substantial effects of classroom, in excess of 50% at some assessment waves. The classroom environment variables significant in the Russian sample were not significant for UK maths classrooms; the only non-performance variable to reach significance was maths anxiety. The results for geography classroom showed very few significant effects. The only significant results were found at time 1 when a modest effect of classroom was shown for student-teacher relations and classroom chaos. Together with the ranking patterns, which were highly consistent for maths classrooms, and largely inconsistent for the geography classrooms, the findings suggest that there may indeed be some degree of informal selection in the Russian schools. It may not necessarily be associated with the enhanced linguistic curricula and ability per se as differences were only observed between the English and Spanish group and English group but not for the English and Chinese group. It may be that parents are influencing the class/teacher allocation process.

To further disentangle effects from teachers and peers that are otherwise confounded in class and teacher groups, relationships between teacher characteristics and measures showing a significant effect of classroom were explored in the Russian sample. Data were available from current subject teachers and the students' primary school teachers. The findings suggested a weak influence of both primary and current subject teachers. Current geography teacher characteristics associated more frequently with classroom measures and these associations persisted across the academic year.

Further investigation showed that teacher characteristics mediated relationships between classroom environment measures and performance. Overall, the mediating models suggested a responsive learning environment for students. For example, maths teacher self-efficacy in student engagement and maths performance at time 1 and time 2 were negatively mediated by maths primary school achievement. This suggests that when primary school achievement was low, teacher self-efficacy was high and in turn, maths performance was high. It may be that teacher self-efficacy was high when they realised the student was learning, which in turn led to better maths performance. This pattern of results may reflect an evocative process whereby the student characteristic elicits behaviour in the teacher, which reinforces teacher self-efficacy. For example, if a student appears to learn concepts quickly, it may lead the teacher to increase the pace of instruction in response, which may lead to higher self-efficacy in the teacher.

Taken together the findings in Chapter 5 suggest a weak influence from primary school years for the Russian sample. However, they do not disentangle specific effects from peers, teachers or selection processes. Variation in peer/teacher dynamics may also be influencing the results. Equally, variation in student characteristics such as ability cannot be discounted.

The findings in Chapter 6 show the influence of variation in student characteristics on differences in performance, differences that are frequently mistaken for an effect of teacher/classroom. In Chapter 5 an effect of classroom was shown in the UK sample for maths anxiety, with high maths anxiety observed for low ability classes. No effect of classroom was observed for maths anxiety in the Russian sample. However, the findings in Chapter 6 revealed that maths anxiety negatively associated with maths performance in both samples, albeit slightly stronger in the UK. Associations also developed differently between the Russian and the UK samples. In the Russian sample, lower prior maths performance negatively associated with future high maths anxiety for Russian students, whereas in the UK sample, higher prior maths anxiety negatively associated with future lower maths performance. Differences across domains were also observed, with the absence of causal relationships between geography anxiety and performance in both samples. It may be that the implementation of the two performance tasks influenced the difference in associations between maths and geography anxiety and performance. The disparity may also result from unequal levels of importance associated with these two academic subjects. The findings show that longitudinal associations between academic anxiety and academic performance manifested differently cross-culturally, and developed differently between academic subjects. Variation found cross-culturally may be a consequence of dissimilarities in education systems.

The focus of Chapter 7 was the dilemma faced by twins' parents of whether their twins should be educated together or separately when they start school. While the majority of schools let parents decide whether to separate their twins, some school boards have a strict policy to teach twins separately. The research into twins' education, however, is limited and suggests inconsistent effects of twins' classroom separation. Consequently parents and educational policymakers are left without clear evidence for educating twins separately or together.

This study sought to address limitations in previous research by comparing school achievement, motivation and cognitive ability in twin pairs taught separately and together in two large representative twin samples in the UK and Quebec (Canada) followed longitudinally from ages 7 to 16 years. The two samples provided the opportunity to take account of two different education systems. Whereas in the UK, most students in secondary education are streamed by ability for their maths classes, in Quebec, there are no formal selection processes and students of mixed ability are taught together.

No sizeable positive or negative average effects of separation were found for twins' achievement, cognitive ability and motivation. This implies negligible effects of being in a particular class with a specific teacher. The few effects found were weak and likely to be explained by education selection processes, such as ability streaming, rather than any real effect of classroom separation.

Limitations

The limitations of each study presented in this thesis are discussed fully within each respective chapter. A more general limitation of the study was with regard to school achievement data collected in the study. Although school achievement is graded by the teacher and aligned very closely with the national curriculums for each sample, there could be issues of generalizability across the samples' countries. For example, in the UK, grades are a product of tests that the teacher has set and marked. Consequently, there may be an element of teacher bias associated with these grades. In Russia, the situation is similar in that the work across the year is set, marked and graded by the teacher to give a final grade. Additionally, not only is there a very narrow range of pass marks (3 to 5), these marks can only be generalized at best, within the sample's school. Because each school employs the same grading system regardless of whether they teach average ability students, students with special educational needs or gifted students, a grade 4 in one school would not necessarily be comparable with the same grade in another school.

Another limitation is the overall sample size of students, classes and teachers. While the sample of students was adequate to test between classrooms, many more schools would be necessary to use a multi-level approach where students were nested in classes, nested in teachers, and nested in schools.

Another issue is with regard to the teacher data collected across both countries. It happened that some teachers decided not to participate, and others did not return their survey booklets. It is unknown whether the sample of participating teachers was biased as result of this attrition. A future study would need to make extra provision for teachers to return any outstanding booklets directly to the researcher. Although teachers were asked to complete the survey during the data collection for students, a few were engaged in other tasks and so the researchers were unable to collect them at the end of the session.

One other issue is the short time frame (across one academic year) that the study covered. While this enabled the exploration of dynamic effects within one academic year, it may have been more informative to also include the last year of primary and second year of secondary education, to give a broader angle as well as the more in-depth one covered here. This may have enabled the teasing out of primary school effects more thoroughly. However, there is a balance to be struck between obtaining more detailed information and duration of the study. In the UK, where data were collected at five assessment points, there was an element of participant fatigue by the fourth and fifth assessments, which led to a small amount of attrition.

One other limitation is the low internal consistency demonstrated by a few of the measures, in particular the homework behaviour and feedback total scale. This is likely due to invariant responses across the items, possibly as a consequence of different homework marking and feedback procedures across schools. Research suggests that Cronbach's alpha is less robust in the face of variation across means, standard deviations and variances (e.g. Dunn et al., 2014; McDonald, 1999). In future analyses of these measures, reliability will be recalculated using omega, an alternative method that is unaffected by invariance across items.

Another limitation of this thesis is the absence of information regarding students' home environments. Factors such as parental education (Chiu et al., 2016), parental occupation (Melhuish et al., 2008) and parental involvement (e.g. Wang & Sheikh-Khalil, 2014) are known to influence academic processes.

Investigating these factors may have provided some insight into potential selection processes and student ability.

Future Directions

Future studies will investigate in another cohort measures that demonstrated a significant effect of classroom/teacher group in this study, to assess whether similar effects are found with a different sample of teachers and students. In the present sample, future analyses will explore across all measures, differences across linguistic groups, as previously only primary school achievement was considered. Further analyses will also be conducted in the UK sample, firstly to take account of the additional assessment waves, and secondly to explore any differences with and without controlling for prior achievement. The data collected from the UK teachers will also be included, to take account of teacher characteristics that may be in influence. For example, to explore, in a streamed environment, potential mediating effects of teacher self-efficacy between classroom measures and performance/achievement. Further analyses will also be conducted to investigate individual differences in maths anxiety and motivational factors. For example, to explore potential associations between maths anxiety and maths self-efficacy/enjoyment, as well as potential associations between maths anxiety and student-teacher relations. Factors that cannot be discounted in the classroom environment.

Conclusions

Teacher/classroom effects appear to be largely elusive and difficult to disentangle from multiple confounding factors. The findings reported in this thesis imply weak influences from current subject teachers and/or previous primary school teachers. Despite investigating numerous aspects of the classroom environment, this thesis is unable to clarify specific effects from such factors as peer influences, selection processes, individual differences in ability and perceptions, teacher characteristics and evocative processes. The findings in this thesis suggest that student outcomes are under multiple influences rather than being predominantly influenced by teacher effects. Overall, the results call for caution in considering ‘added value’ or ‘teacher effect’ measures as valid criteria for current education policies that affect teacher promotion and employment prospects.

References

Adler, N., & Stewart, J. (2007). The MacArthur scale of subjective social status. *MacArthur Research Network on SES & Health. Retrieved from http://www. macses. ucsf. edu/Research/Psychosocial/subjective. php*.

Anderman, E. M., Maehr, M. L., & Midgley, C. (1999). Declining motivation after the transition to middle school: Schools can make a difference. *Journal of Research & Development in Education*.

Asbury, K., Almeida, D., Hibel, J., Harlaar, N., &Plomin, R. (2008). Clones in the classroom: A daily diary study of the non-shared environmental relationship between monozygotic twin differences in school experience and achievement. *Twin Research and Human Genetics*, *11*(06), 586-595.

Ashcraft, M. H., & Krause, J. A. (2007). Working memory, math performance, and math anxiety. *Psychonomic bulletin & review*, *14*(2), 243-248.

Ashcraft, M. H., & Moore, A. M. (2009). Mathematics anxiety and the affective drop in performance. *Journal of Psychoeducational Assessment*, *27*(3), 197-205.

Balzer, W. K., Kihm, J. A., Smith, P. C., Irwin, J. L., Bachiochi, P. D., Robie, C., ... & Parra, L. F. (2000). Users‟ Manual for the Job Descriptive Index (JDI; 1997 version) and the Job in General Scales. Electronic resources for the JDI and JIG. Bowling Green, OH: Bowling Green State University.

Baumert, J., Kunter, M., Blum, W., Brunner, M., Voss, T., Jordan, A., ... & Tsai, Y. M. (2010).Teachers' mathematical knowledge, cognitive activation in the classroom, and student progress. *American Educational Research Journal*, *47*(1), 133-180.

Beilock, S. L., Gunderson, E. A., Ramirez, G., & Levine, S. C. (2010). Female teachers’ math anxiety affects girls’ math achievement. *Proceedings of the National Academy of Sciences*, *107*(5), 1860-1863.

Bentler, P. M., & Bonett, D. G. (1980). Significance tests and goodness of fit in the analysis of covariance structures. *Psychological bulletin*, *88*(3), 588.

Blatchford, P., Bassett, P., & Brown, P. (2011). Examining the effect of class size on classroom engagement and teacher–pupil interaction: Differences in relation to pupil prior attainment and primary vs. secondary schools. *Learning and Instruction*, *21*(6), 715-730.

Blatchford, P., Bassett, P., Goldstein, H., & Martin, C. (2003). Are class size differences related to pupils' educational progress and classroom processes? Findings from the institute of education class size study of children aged 5–7 years. *British Educational Research Journal*, *29*(5), 709-730.

Boivin, M., Brendgen, M., Dionne, G., Dubois, L., Pérusse, D., Robaey, P., ... & Vitaro, F. (2013). The Quebec newborn twin study into adolescence: 15 years later. *Twin Research and Human Genetics*, *16*(01), 64-69.

Bong, M., & Skaalvik, E. M. (2003). Academic self-concept and self-efficacy: How different are they really? *Educational Psychology Review*, *15*(1), 1-40.

Bosworth, R. (2014). Class size, class composition, and the distribution of student achievement. *Education Economics*, *22*(2), 141-165.

Brackett, M. A., Mayer, J. D., & Warner, R. M. (2004). Emotional intelligence and its relation to everyday behaviour. *Personality and Individual differences*, *36*(6), 1387-1402.

Bray, M., & Kwo, O. (2013). Behind the façade of fee-free education: Shadow education and its implications for social justice. *Oxford Review of*

Education, *39*(4), 480-497.

Brühwiler, C., & Blatchford, P. (2011). Effects of class size and adaptive teaching competency on classroom processes and academic outcome. *Learning and Instruction*, *21*(1), 95-108.

Burgess, S. M., Dickson, M., & Macmillan, L. (2014). Selective schooling systems increase inequality. *The Centre for the Analysis of Social Policy University of Bath*, *.UK.* Retrieved from: http://www.bath.ac.uk/casp

Burke, M. A., & Sass, T. R. (2013). Classroom peer effects and student achievement. *Journal of Labor Economics*, *31*(1), 51-82.

Byrne, B. M. (2012). *Structural equation modeling with Mplus: Basic concepts, applications and programming.* Hove, Sussex: Routledge.

Byrne, B., Coventry, W. L., Olson, R. K., Wadsworth, S. J., Samuelsson, S., Petrill, S. A., Willcutt, E. G., & Corley, R. (2010). “Teacher effects” in early literacy development: Evidence from a study of twins. *Journal of Educational Psychology*, *102*(1), 32.

Caprara, G. V., Barbaranelli, C., Steca, P., & Malone, P. S. (2006). Teachers' self-efficacy beliefs as determinants of job satisfaction and students' academic achievement: A study at the school level. *Journal of school psychology*, *44*(6), 473-490.

Carman, K. G., & Zhang, L. (2012). Classroom peer effects and academic achievement: Evidence from a Chinese middle school. *China Economic Review*, *23*(2), 223-237.

CAT Scores (Cognitive Abilities Test, Third Edition (CAT3)). http://www.gl-assessment.co.uk/products/cat-cognitive-abilities-test/cat-cognitive-abilities-test-faqs

Caughy, M.O.B, DiPietro, J.A., & Strobino, D.M. (1994). Day-care participation as a protective factor in the cognitive development of low-income children. *Child Development, 65*(2) 457–71.

Chen, J., Li, X., Zhang, J., Natsuaki, M. N., Leve, L. D., Harold, G. T., ... & Ge, X. (2013). The Beijing Twin Study (BeTwiSt): A longitudinal study of child and adolescent development. *Twin Research and Human Genetics*, *16*(01), 91-97.

Cherkas, L. (2015). Results of questionnaire survey of parents of young twins regarding placement of twins in primary school. *TAMBA.* Retrieved from:https://www.tamba.org.uk/Reports 1/3/15.

Chetty, R., Friedman, J. N., & Rockoff, J. E. (2011). The long-term impacts of teachers: Teacher value-added and student outcomes in adulthood (No. w17699). *National Bureau of Economic Research.*

Chiu, J., Economos, J., Markson, C., Raicovi, V., Howell, C., Morote, E. S., & Inserra, A. (2016). Which matters most? Perceptions of family income or parental education on academic achievement. *New York Journal of Student Affairs*, *16*(2), 3-16.

Chmielewski, A. K., Dumont, H., & Trautwein, U. (2013). Tracking effects depend on tracking type an international comparison of students' mathematics self-concept. *American Educational Research Journal*, 0002831213489843.

Christodoulo, D. (2014). *Seven myths about education.* Oxon: Routledge.

Cobb, B. (2004). Assessing job satisfaction and emotional intelligence in public school teachers *Doctoral dissertation, Western Kentucky University.*

Cooper, S. E., & Robinson, D. A. (1991). The relationship of mathematics self-efficacy beliefs to mathematics anxiety and performance. *Measurement and Evaluation in Counseling and Development*.

Coughlan, S. (2016, December 6). Pisa tests: UK lags behind in global school rankings. *BBC News*. http://www.bbc.co.uk/news/education-38157811

Coventry, W. L., Byrne, B., Coleman, M., Olson, R. K., Corley, R., Willcutt, E., & Samuelsson, S. (2009). Does classroom separation affect twins' reading ability in the early years of school? *Twin Research and Human Genetics*, *12*(05), 455-461.

Dearing, E., McCartney, K., & Taylor, B.A. (2009). Does higher quality early child-care promote low-income children's math and reading achievement in middle childhood? *Child Development, 80*(5) 1329–49.

Deary, I. J., Strand, S., Smith, P., & Fernandes, C. (2007). Intelligence and educational achievement. *Intelligence*, *35*(1), 13-21.

Devine, A., Fawcett, K., Szűcs, D., & Dowker, A. (2012). Gender differences in mathematics anxiety and the relation to mathematics performance while controlling for test anxiety. *Behavioral and brain functions*, *8*(1), 33.

Dicke, T., Parker, P. D., Marsh, H. W., Kunter, M., Schmeck, A., & Leutner, D. (2014). Self-efficacy in classroom management, classroom disturbances, and emotional exhaustion: A moderated mediation analysis of teacher candidates. *Journal of Educational Psychology*, *106*(2), 569.

Ding, W., & Lehrer, S. F. (2007). Do peers affect student achievement in China's secondary schools? *The Review of Economics and Statistics*, *89*(2), 300-312.

Dunn, T. J., Baguley, T., & Brunsden, V. (2014). From alpha to omega: A practical solution to the pervasive problem of internal consistency estimation. *British Journal of Psychology, 105*(3), 399-412.

Dweck, C. S., Chiu, C. Y., & Hong, Y. Y. (1995). Implicit theories and their role in judgments and reactions: A word from two perspectives. *Psychological*

Inquiry, 6(4), 267-285.

Ecalle, J., Magnan, A., &Gibert, F. (2006). Class size effects on literacy skills and literacy interest in first grade: A large-scale investigation. *Journal of School Psychology*, *44*(3), 191-209.

Eccles, J. (2012). Expectancy-value theory and gendered academic and occupational choices. In annual meeting of the American Educational Research Association, Vancouver, British Columbia, Canada.

Eccles, J. S. (1999). The development of children ages 6 to 14. *The future of children*, 30-44.

Eccles, J. S., Midgley, C., Wigfield, A., Buchanan, C. M., Reuman, D., Flanagan, C., & Mac Iver, D. (1993). Development during adolescence: The impact of stage-environment fit on young adolescents' experiences in schools and in families. *American psychologist*, *48*(2), 90.

Field, A. (2011). *Discovering statistics using IBM SPSS statistics*. Sage.

Figlio, D. N., & Lucas, M. E. (2000). *What's in a grade? School report cards and house prices* (No. w8019). National Bureau of Economic Research.

Geiser, C. (2013). *Data analysis with MPlus*. English Language Edition, the Guilford Press, New York.

Geoffroy, M.C., Côté, S.M., Borge, A.I., Larouche, F., Séguin, J.R., &Rutter, M. (2007). Association between non-maternal care in the first year of life and children’s receptive language skills prior to school entry: The moderating role of socioeconomic status. *Journal of Child Psychology and Psychiatry, 48(*5):490–97.

Gibbons, S., Machin, S., & Silva, O. (2013). Valuing school quality using boundary discontinuities. *Journal of Urban Economics*, *75*, 15-28.

Gliner, G. S. (1987). The relationship between mathematics anxiety and achievement variables. *School Science and mathematics*, *87*(2), 81-87.

Goddard, R. D., Hoy, W. K., & Hoy, A. W. (2000). Collective teacher efficacy: Its meaning, measure, and impact on student achievement. *American Educational Research Journal*, *37*(2), 479-507.

Goetz, T., Bieg, M., Lüdtke, O., Pekrun, R., & Hall, N. C. (2013). Do girls really experience more anxiety in mathematics? *Psychological science*, *24*(10), 2079-2087.

Goldstein, H., & Blatchford, P. (1998). Class size and educational achievement: A review of methodology with particular reference to study design. *British Educational Research Journal*, *24*(3), 255-268.

Guyon, N., Maurin, E., & McNally, S. (2012). The effect of tracking students by ability into different schools: a natural experiment. *Journal of Human Resources*, *47*(3), 684-721.

Hamre, B. K., & Pianta, R. C. (2005). Can instructional and emotional support in the first- grade classroom make a difference for children at risk of school failure? *Child development*, *76*(5), 949-967.

Hanushek, E. A. (1999). Some findings from an independent investigation of the Tennessee STAR experiment and from other investigations of class size effects. *Educational Evaluation and Policy Analysis*, *21*(2), 143-163.

Hanushek, E. A. (2011). The economic value of higher teacher quality. *Economics of Education Review*, *30*(3), 466-479.

Harlaar, N., Santtila, P., Björklund, J., Alanko, K., Jern, P., Varjonen, M., Von der Pahlen, B.,& Sandnabba, K. (2008). Retrospective reports of parental physical affection and parenting style: a study of Finnish twins. *Journal of Family Psychology*, *22*(4), 605.

Hart, S. A., Logan, J. A., Soden-Hensler, B., Kershaw, S., Taylor, J., &

Schatschneider, C. (2013). Exploring how nature and nurture affect the development of reading: An analysis of the Florida Twin Project on Reading. *Developmental psychology*, *49*(10), 1971.

Hattie, J. A. (2002). Classroom composition and peer effects. *International Journal of Educational Research*, *37*(5), 449-481.

Haworth, C. M. A., Asbury, K., Dale, P. S., Plomin, R. (2011) Added Value Measures in Education Show Genetic as Well as Environmental Influence. *PLOS ONE 6*(2): e16006. doi:10.1371/journal.pone.0016006

Haworth, C. M., Davis, O. S., & Plomin, R. (2013). Twins Early Development Study (TEDS): a genetically sensitive investigation of cognitive and behavioral development from childhood to young adulthood. *Twin Research and Human Genetics*, *16*(01), 117-125.

Haworth, C. M., Davis, O. S., Hanscombe, K. B., Kovas, Y., Dale, P. S., & Plomin, R. (2013). Understanding the science-learning environment: A genetically sensitive approach. *Learning and individual differences*, *23*, 145-150.

Hembree, R. (1990). The nature, effects and relief of mathematics anxiety. *Journal for Research in Mathematics Education*, 21, 33-46.

Hopko, D. R., Mahadevan, R., Bare, R. L., & Hunt, M. K. (2003). The abbreviated math anxiety scale (AMAS): Construction, validity, and reliability. *Assessment, 10*(2), 178–182. doi: 10.1177/1073191103010002008

Hopko, D. R., McNeil, D. W., Zvolensky, M. J., & Eifert, G. H. (2002). The relation between anxiety and skill in performance-based anxiety disorders: A behavioral formulation of social phobia. *Behavior Therapy, 32*(1), 185-207.

Horwitz, E. (2001). Language anxiety and achievement. *Annual Review of Applied Linguistics, 21,* 112-126.

Hughes, J. N., Im, M. H., & Wehrly, S. E. (2014). Effect of peer nominations of teacher–student support at individual and classroom levels on social and academic outcomes. *Journal of School Psychology, 52*(3), 309-322.

Hund, A. M., & Minarik, J. L. (2006). Getting from here to there: Spatial anxiety, way-finding strategies, direction type, and wayfinding efficiency. *Spatial Cognition and Computation, 6*(3), 179-201.

Jansen, B. R., Louwerse, J., Straatemeier, M., Van der Ven, S. H., Klinkenberg, S., & Van der Maas, H. L. (2013). The influence of experiencing success in math on math anxiety, perceived math competence, and math performance. *Learning and Individual Differences, 24,* 190-197.

Jöreskog, K. G. (1971). Simultaneous factor analysis in several populations. *Psychometrika, 36*(4), 409-426.

Kang, C. (2007). Classroom peer effects and academic achievement: Quasi-randomization evidence from South Korea. *Journal of Urban Economics, 61*(3), 458-495.

Kaplan, E., Fein, D., Kramer, J., Delis, D., & Morris, R. (1999). *WISC-III as a Process Instrument (WISC-III-PI).* New York: The Psychological Corporation.

Kelly, S., & Carbonaro, W. (2012). Curriculum tracking and teacher expectations: Evidence from discrepant course taking models. *Social Psychology of Education, 15*(3), 271-294.

Konstantopoulos, S. (2008). Do small classes reduce the achievement gap between low and high achievers? Evidence from Project STAR. *The Elementary School Journal, 108*(4), 275-291.

Kovas, Y., & Plomin, R. (2006). Generalist genes: implications for the cognitive sciences. Trends in cognitive sciences, 10(5), 198-203.

Kovas, Y., Garon-Carrier, G., Boivin, M., Petrill, S. A., Plomin, R., Malykh, S. B., ... & Brendgen, M. (2015). Why children differ in motivation to learn: Insights from over 13,000 twins from 6 countries. *Personality and Individual Differences, 80*, 51-63.

Kovas, Y., Harlaar, N., Petrill, S. A., & Plomin, R. (2005). 'Generalist genes' and mathematics in 7-year-old twins. *Intelligence, 33*(5), 473-489.

Kovas, Y., Haworth, C. M., Dale, P. S., Plomin, R., Weinberg, R. A., Thomson, J. M., & Fischer, K. W. (2007). The genetic and environmental origins of learning abilities and disabilities in the early school years. *Monographs of the Society for Research in Child Development,* i-156.

Kovas, Y., Malykh, S., Gaysina, D. (Eds.) (2016). *Behavioural Genetics for Education.* Palgrave Macmillan UK. ISBN 978-1-137-43732-7

Kovas, Y., Petrill, S.A., Plomin, R. (2007). The Origins of Diverse of Mathematics: Generalist genes but Specialist Environments. *Journal of Educational Psychology, 99*(1), 128-139.

Krapohl, E., Rimfeld, K., Shakeshaft, N. G., Trzaskowski, M., McMillan, A., Pingault, J. B., ... & Plomin, R. (2014). The high heritability of educational achievement reflects many genetically influenced traits, not just intelligence. *Proceedings of the National Academy of Sciences, 111*(42), 15273-15278.

Krinzinger, H., Kaufmann, L., & Willmes, K. (2009). Math anxiety and math ability in early primary school years. *Journal of Psychoeducational Assessment, 27*(3), 206-225.

Lalonde, N. & Moisan, D. (2003). Les jumeaux à l'école: dans la meme classe ou dans des classes séparées… qui doit décider?

Lavy, V., & Schlosser, A. (2011). Mechanisms and impacts of gender peer effects at school. *American Economic Journal: Applied Economics, 3*(2), 1-33.

Lee, J. (2009). Universals and specifics of math self-concept, math self-efficacy, and math anxiety across 41 PISA 2003 participating countries. *Learning and Individual Differences, 19*(3), 355-365.

Lee, S., Turner, L. J., Woo, S., & Kim, K. (2014). All or nothing? The impact of school and classroom gender composition on effort and academic achievement (No. w20722). *National Bureau of Economic Research*.

Loeb, S., Soland, J., & Fox, L. (2014). Is a good teacher a good teacher for all? Comparing value-added of teachers with their English learners and non-English learners. *Educational Evaluation and Policy Analysis, 36*(4), 457-475.

Luo, Y. L., Haworth, C. M., & Plomin, R. (2010). A novel approach to genetic and environmental analysis of cross-lagged associations over time: The cross-lagged relationship between self-perceived abilities and school achievement is mediated by genes as well as the environment. *Twin Research and Human Genetics, 13*(05), 426-436.

Ma, X., & Xu, J. (2004). The causal ordering of mathematics anxiety and mathematics achievement: A longitudinal panel analysis. *Journal of Adolescence, 27,* 165- 179.

Maaz, K., Trautwein, U., Lüdtke, O., & Baumert, J. (2008). Educational transitions and differential learning environments: How explicit between-school tracking contributes to social inequality in educational outcomes. *Child Development Perspectives, 2*(2), 99-106.

Malanchini, M., Rimfeld, K., Shakeshaft, N. G., Rodic, M., Schofield, K., Selzam, S., ... & Kovas, Y. (2017). The genetic and environmental aetiology of spatial, mathematics and general anxiety. *Scientific Reports*, *7*. doi: 10.1038/srep42218

Maloney, E. A., Risko, E. F., Ansari, D., & Fugelsang, J. (2010). Mathematics anxiety affects counting but not subitizing during visual enumeration. *Cognition, 114*(2), 293-297.

Marsh, H. W. (1987). The big-fish-little-pond effect on academic self-concept. *Journal of Educational Psychology, 79*(3), 280.

Marsh, H. W., Kong, C. K., & Hau, K. T. (2000). Longitudinal multilevel models of the big-fish-little-pond effect on academic self-concept: counterbalancing contrast and reflected-glory effects in Hong Kong schools. *Journal of Personality and Social Psychology, 78*(2), 337.

Marsh, H. W., Martin, A. J., & Cheng, J. (2008). How we judge ourselves from different perspectives: Contextual influences on self-concept formation. Advances in motivation and achievement: *Social Psychological Perspectives, 15,* 315-356.

Maslach, C., Jackson, S. E., & Leiter, M. P. (1996). The Maslach Burnout Inventory. (3rd ed.). Palo Alto, CA: Consulting Psychologists Press.

Matheny, A.P., Wachs, T.D., Ludwig, J.L., & Phillips, K. (1995), Bringing order out of chaos: psychometric characteristics of the confusion hubbub and order scale, *Journal of Applied Developmental Psychology, 16,* 429-444.

Maulana, R., Opdenakker, M. C., & Bosker, R. (2014). Teacher–student interpersonal relationships do change and affect academic motivation: A multilevel growth curve modelling. *British Journal of Educational Psychology, 84*(3), 459-482.

Maulana, R., Opdenakker, M. C., Stroet, K., & Bosker, R. (2013). Changes in teachers' involvement versus rejection and links with academic motivation during the first year of secondary education: A multilevel growth curve analysis. *Journal of Youth and Adolescence, 42*(9), 1348-1371.

McCarthy, D. (1972). *McCarthy Scales of Children's Abilities.* New York: The Psychological Corporation.

McDonald, R. (1999). The dimensionality of tests and items. *British Journal of Mathematical and Statistical Psychology, 34,* 100-117.

McFarland, D. A. (2001). Student resistance: How the formal and informal organization of classrooms facilitate everyday forms of student defiance 1. *American Journal of Sociology, 107*(3), 612-678.

Melhuish, E. C., Phan, M. B., Sylva, K., Sammons, P., Siraj- Blatchford, I., & Taggart, B. (2008). Effects of the home learning environment and preschool center experience upon literacy and numeracy development in early primary school. *Journal of Social Issues, 64*(1), 95-114.

Middlemass, J. (2014). National curriculum assessments at key stage 2 in England, 2014 (Provisional). *Department for Education*. Retrieved from:https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/347653/SFR30_2014_Text.pdf

Midgley, C., Eccles, J. S., & Feldlaufer, H. (1991). Classroom environment and the transition to junior high school. In B. J. Fraser and H.J. Walberg (Eds.), *Educational Environments: Evaluation, Antecedents and Consequences* (pp. 113-139). Oxford: Pergamon Press.

Midgley, C., Feldlaufer, H., & Eccles, J. S. (1989). Student/teacher relations and attitudes toward mathematics before and after the transition to junior high school. *Child Development*, 981-992.

Morony, S., Kleitman, S., Lee, Y. P., & Stankov, L. (2013). Predicting achievement: Confidence vs. self-efficacy, anxiety, and self-concept in Confucian and European countries. *International Journal of Educational Research, 58,* 79-96.

Mullis, I. V. S., Martin, M. O., Foy, p., & Hooper, M. (2016). *TIMSS 2015 International Results in Science.* Retrieved from Boston College, TIMSS & PIRLS International Study Center website: http://timssandpirls.bc.edu/timss2015/international-results/

Mullis, I. V., Martin, M. O., Foy, P., & Arora, A. (2012). *TIMSS 2011 international results in mathematics.* International. Association for the Evaluation of Educational Achievement. Herengracht 487, Amsterdam, 1017 BT, The Netherlands.

Murayama, K., Pekrun, R., Lichtenfeld, S., & VomHofe, R. (2012). Predicting long‐ term growth in students' mathematics achievement: The unique contributions of motivation and cognitive strategies. *Child Development, 84*(4), 1475-1490.

Murphy, M. M., & Mazzocco, M. M. M (2008). Mathematics learning disabilities in girls with Fragile X or Turner syndrome during late elementary school. *Journal of Learning Disabilities, 41*(1), 29-46.

Muthén LK, Muthén BO. (1998-2015). *Mplus User's Guide* (7th Ed). Los Angeles, CA: Muthén & Muthén

NAEP (2005). *Mathematics: Student background questionnaire (Grade 4*). Retrieved June 8, 2006, from nces.ed.gov

NICARM (2016). *National Information Centre On Academic Recognition and Mobility*. Retrieved from: http://www.russianenic.ru/english/rus/#10

Nye, B. (2000). The effects of small classes on academic achievement: The results of the Tennessee class size experiment. *American Educational Research Journal, 37*(1), 123–151. Retrieved from http://aer.sagepub.com/content/37/1/123.short

Nye, B., Konstantopoulos, S., & Hedges, L. V. (2004). How large are teacher effects? *Educational Evaluation and Policy Analysis, 26*(3), 237–257. doi:10.3102/01623737026003237

O'Dwyer, L. M., Wang, Y., & Shields, K. A. (2015). Teaching for conceptual understanding: A cross-national comparison of the relationship between teachers' instructional practices and student achievement in mathematics. *Large-scale Assessments in Education, 3*(1), 1.

OECD (2010). PISA 2009 Results: What Students Know and Can Do – Student Performance in Reading, Mathematics and Science (Volume I). Retrieved fromhttp://dx.doi.org/10.1787/9789264091450-en

OECD (2013). *PISA 2012 Results in Focus: What 15-year-olds know and what they can do with what they know: Key results from PISA 2012.* Retrieved from http://www.oecd.org/pisa/keyfindings/pisa-2012-results.htm

OECD (2016). *PISA 2015 Results in Focus: (Volume I), Excellence and Equity in Education.* Retrieved from https://www.oecd.org/pisa/pisa-2015-results-in-focus.pdf

OECD (2016). *PISA 2015 Results in Focus: (Volume I), Excellence and Equity in Education.* Retrieved from https://www.oecd.org/pisa/pisa-2015-results-in-focus.pdf

OECD (2016). *PISA 2015 Results in Focus: (Volume II), Policies and Practices for successful Schools.* Retrieved from https://www.oecd.org/pisa/pisa-2015-results-in-focus.pdf

Office for National Statistics (2014). *Births in England and Wales by characteristics of birth, 2, 2015.* Retrieved from: http://www.ons.gov.uk/ons/rel/vsob1/characteristics-of-birth-2--england-and-wales/2013/sb-characteristics-of-birth-2.html 15/02/2016

Olson, R. K., Keenan, J. M., Byrne, B., & Samuelsson, S. (2014). Why do children differ in their development of reading and related skills? *Scientific Studies of Reading, 18*(1), 38-54.

Opdenakker, M. C., & Damme, J. (2001). Relationship between school composition and characteristics of school process and their effect on mathematics achievement. *British Educational Research Journal, 27*(4), 407-432.

Pajares, F., & Miller, M. D. (1994). Role of self-efficacy and self-concept beliefs in mathematical problem solving: A path analysis. *Journal of Educational Psychology,* 86(2), 193.

Petrides, K. V., & Furnham, A. (2004). *Technical manual of the Trait Emotional Intelligence Questionnaire (TEIQue).* London: University of London, Institute of Education.

Petrides, K. V., Pérez, J. C., & Furnham, A. (2003). The Trait Emotional Intelligence Questionnaire (TEIQue). A measure of emotional self-efficacy. *In XI Biennial Meeting of the International Society for the Study of the Individual Differences.*

Petrill, S. A., Kovas, Y., Hart, S. A., Thompson, L. A., &Plomin, R. (2009). The genetic and environmental etiology of high math performance in 10-year-old twins. *Behavior Genetics, 39*(4), 371-379.

Pianta, R. C., Belsky, J., Vandergrift, N., Houts, R., & Morrison, F. J. (2008).Classroom effects on children’s achievement trajectories in elementary school. *American Educational Research Journal, 45*(2), 365-397.

Pianta, R. C., Steinberg, M. S., & Rollins, K. B. (1995). The first two years of school: Teacher-child relationships and deflections in children's classroom adjustment. *Development and Psychopathology, 7*(02), 295-312.

PISA (OECD Programme for International Student Assessment): www.pisa.oecd.org.

Plomin, R., & Bergeman, C. S. (1991). The nature of nurture: Genetic influences on “environmental” measures. *Behavioral and Brain Sciences, 14,* 373–427.

Polderman, T. J., Bartels, M., Verhulst, F. C., Huizink, A. C., Van Beijsterveldt, C. E., & Boomsma, D. I. (2010). No effect of classroom sharing on educational achievement in twins: a prospective, longitudinal cohort study. *Journal of Epidemiology and Community Health, 64*(01), 36-40.

Price, T. S., Freeman, B., Craig, I., Petrill, S. A., Ebersole, L., & Plomin, R. (2000). Infant zygosity can be assigned by parental report questionnaire data. *Twin Research, 3*(03), 129-133.

Ramirez, G., Gunderson, E. A., Levine, S. C., & Beilock, S. L. (2013). Math anxiety, working memory, and math achievement in early elementary school. *Journal of Cognition and Development, 14*(2), 187-202.

Raven, J.C., Court, J.H., & Raven, J. (1996). *Manual for Raven's Progressive Matrices and Vocabulary Scales.* Oxford: Oxford University Press.

Reiss, D., Hetherington, E. M., Plomin, R., Howe, G. W., Simmens, S. J., Henderson, S. H., ... & Law, T. (1995). Genetic questions for environmental studies: Differential parenting and psychopathology in adolescence. *Archives of General Psychiatry, 52*(11), 925-936.

Reyes, M. R., Brackett, M. A., Rivers, S. E., White, M., & Salovey, P. (2012). Classroom emotional climate, student engagement, and academic achievement. *Journal of Educational Psychology, 104*(3), 700.

Rimfeld, K., Kovas, Y., Dale, P. S., &Plomin, R. (2015). Pleiotropy across academic subjects at the end of compulsory education. *Scientific Reports, 5*.

Rodic, M., Zhou, X., Tikhomirova, T., Wei, W., Malykh, S., Ismatulina, V., Sabirova, E., Davidova, Y., Tosto, M., Lemelin, J-P., & Kovas, Y. (2015). Cross- cultural investigation into cognitive underpinnings of individual differences in early arithmetic. *Developmental Science, 18*(1), 165-174.

Rowan, B., Chiang, F. S., & Miller, R. J. (1997). Using research on employees' performance to study the effects of teachers on students' achievement. *Sociology of Education*, 256-284.

Rowan, B., Correnti, R., & Miller, R. (2002). What large-scale survey research tells us about teacher effects on student achievement: Insights from the Prospects study of elementary schools*. The Teachers College Record, 104*(8), 1525-1567.

Saigh, P. A., & Khouri, A. (1983). The concurrent validity of the mathematics anxiety rating scale for adolescents (MARS-A) in relation to the academic achievement of Lebanese students. *Educational and Psychological Measurement, 43*(2), 633-637.

Samuelson, P. N. (2004). *Economics* (18th ed.) Toronto: McGraw-Hill/Irwin.

SAT scores http://www.education.gov.uk/schools/teachingandlearning/assessment/keystage2/a00203532/about-ks2

Segal, N. L., & Russell, J. M. (1992). Twins in the classroom: School policy issues and recommendations. *Journal of Educational and Psychological Consultation, 3*(1), 69-84.

Sharrow, D. J., & Anderson, J. J. (2016). A Twin Protection Effect? Explaining Twin Survival Advantages with a Two-Process Mortality Model. *PloS One, 11*(5), e0154774.

Shavelson, R. J., Hubner, J. J., & Stanton, G. C. (1976). Self-concept: Validation of construct interpretations. *Review of Educational Research, 46(*3), 407-441.

Siegler, R. S. & Opfer, J. E. (2006). Representational change and children's numerical estimation. *Cognitive Psychology*. doi:10.1016/j.cogpsych.2006.09.00

Skaalvik, E. M., & Skaalvik, S. (2007). Dimensions of teacher self-efficacy and relations with strain factors, perceived collective teacher efficacy, and teacher burnout. *Journal of Educational Psychology, 99*(3), 611.

Smith, P., Fernandes, C., & Strand, S. (2001). *Cognitive abilities test 3 (CAT3).* Windsor, England: nferNelson.

Spinath, B., Spinath, F. M., Harlaar, N., & Plomin, R. (2006). Predicting school achievement from general cognitive ability, self-perceived ability, and intrinsic value. *Intelligence, 34*(4), 363-374.

Spinath, F. M., Spinath, B., & Plomin, R. (2008). The nature and nurture of intelligence and motivation in the origins of sex differences in elementary school achievement. *European Journal of Personality, 22*(3), 211-229.

Stanovich, K. E. (1986). Matthew effects in reading: Some consequences of individual differences in the acquisition of literacy. *Reading Research Quarterly, 360*-407.

TAMBA. (2009). *Mulitple choice: The educational needs and experiences of multiple birth children.* Retrieved from The Twin and Multiple Birth Association website: https://www.tamba.org.uk/parenting

TAMBA. (2010). *Part III. Primary and secondary schools admissions and appeals pack: Information guide for parents and professionals in England.* Retrieved from http://www.nhs.uk/Conditions/pregnancy-and-baby/Pages/twins-at-school.aspx

Taylor, B. A., & Fraser, B. J. (2013). Relationships between learning environment and mathematics anxiety. *Learning Environments Research, 16(*2), 297-313.

Taylor, J., Roehrig, A. D., Hensler, B. S., Connor, C. M., & Schatschneider, C. (2010). Teacher quality moderates the genetic effects on early reading. *Science, 328*(5977), 512-514.

Thatcher, J. B., & Perrewe, P. L. (2002). An empirical examination of individual traits as antecedents to computer anxiety and computer self-efficacy. *MIS Quarterly,* 381-396.

Torgesen, J. K., Wagner, R. K., & Rashotte, C. A. (1999). *TOWRE (Test of Word Reading Efficiency).* Austin, Texas: Pro-Ed

Trautwein, U., Lüdtke, O., Marsh, H. W., Köller, O., & Baumert, J. (2006). Tracking, grading, and student motivation: Using group composition and status to predict self-concept and interest in ninth-grade mathematics. *Journal of Educational Psychology, 98*(4), 788.

Tschannen-Moran, M., & Hoy, A. W. (2001). Teacher efficacy: Capturing an elusive construct. *Teaching and Teacher Education, 17*(7), 783-805.

Tschannen-Moran, M., & Hoy, A. W. (2007). The differential antecedents of self-efficacy beliefs of novice and experienced teachers*. Teaching and*

Teacher Education, 23(6), 944-956.

Tschannen-Moran, M., Hoy, A. W., & Hoy, W. K. (1998). Teacher efficacy: Its meaning and measure. *Review of Educational Research, 68*(2), 202-248.

Tully, L. A., Moffitt, T. E., Caspi, A., Taylor, A., Kiernan, H., & Andreou, P. (2004). What effect does classroom separation have on twins' behavior, progress at school, and reading abilities? *Twin Research, 7*(02), 115-124.

Van Leeuwen, M., Van Den Berg, S. M., van Beijsterveldt, T. C., & Boomsma, D. I. (2005). Effects of twin separation in primary school. *Twin Research and Human Genetics, 8*(04), 384-391.

Vukovic, R. K., Kieffer, M. J., Bailey, S. P., & Harari, R. R. (2013). Mathematics anxiety in young children: Concurrent and longitudinal associations with mathematical performance. *Contemporary Educational Psychology, 38*(1), 1-10.

Vukovic, R. K., Roberts, S. O., & Green Wright, L. (2013). From parental involvement to children's mathematical performance: The role of mathematics anxiety. *Early Education & Development, 24*(4), 446-467.

Wai, J., Lubinski, D., & Benbow, C. P. (2009). Spatial ability for STEM domains: Aligning over 50 years of cumulative psychological knowledge solidifies its importance. *Journal of Educational Psychology, 101*(4), 817–835.

Wang, M. T., & Eccles, J. S. (2012). Social support matters: Longitudinal effects of social support on three dimensions of school engagement from middle to high school. *Child Development, 83*(3), 877-895.

Wang, M. T., & Holcombe, R. (2010). Adolescents' perceptions of school environment, engagement, and academic achievement in middle school. *American Educational Research Journal, 47*(3), 633-662.

Wang, M. T., & Sheikh‐ Khalil, S. (2014). Does parental involvement matter for student achievement and mental health in high school? *Child Development, 85*(2), 610-625.

Wang, Z., Hart, S. A., Kovas, Y., Lukowski, S., Soden, B., Thompson, L. A., ... Petrill, S. A. (2014). 'Who is afraid of math? Two sources of genetic variance for mathematical anxiety'. *Journal of Child Psychology and Psychiatry, 55*(9), 1056-1064

Webbink, D., Hay, D., & Visscher, P. M. (2007). Does sharing the same class in school improve cognitive abilities of twins? *Twin Research and Human Genetics, 10*(04), 573-580.

Wechsler, D. (1992). *Wechsler Intelligence Scale for Children (3rd Edition) UK (WISC-III-UK) Manual.* London: The Psychological Corporation.

Woessmann, L. (2016). The importance of school systems: Evidence from international differences in student achievement. *The Journal of Economic Perspectives, 30*(3), 3-31.

Wu, S., Amin, H., Barth, M., Malcarne, V., & Menon, V. (2012).Math anxiety in second and third graders and its relation to mathematics achievement. *Frontiers in Psychology, 3*, 162

Xu, J., & Hampden-Thompson, G. (2012). Cultural reproduction, cultural mobility, cultural resources, or trivial effect? A comparative approach to cultural capital and educational performance. *Comparative Education Review, 56(*1), 98-124.

Zuffianò, A., Alessandri, G., Gerbino, M., Kanacri, B. P. L., Di Giunta, L., Milioni, M., & Caprara, G. V. (2013). Academic achievement: The unique contribution of self-efficacy beliefs in self-regulated learning beyond intelligence, personality traits, and self-esteem. *Learning and Individual*

Differences, *23,* 158-162.

Appendices

Appendix 1.

Internal validity of student maths and geography classroom measures

The results presented in Tables 1.1 to 1.4 show that the majority of measures demonstrated acceptable internal validity for both samples. There were a few exceptions that revealed very low Cronbach's alpha. In both the UK and Russian samples, homework behaviour (alpha =.149 to .551), and homework total scale (alpha = .430 to .646) were very low. This may reflect the items, for example, *'I do my homework while watching television'*, may not be indicative of the way students do their homework. Many UK schools use software packages for maths which means that often students have to complete their homework at the school. Low Cronbach's alpha was also observed for maths classroom environment in the UK at time 1 (alpha = .598). This may be due to students having had little classroom experience by the first testing date at the start of the school year. The Cronbach alpha improved considerably by the end of the study at time 5 (alpha = .814).

The results of the Cronbach's alpha may indicate that this assessment of internal reliability is not appropriate for these data, particularly when there may be variation in responses due to external factors. Researchers have suggested that certain assumptions apply when using alpha and these are frequently not met during psychological testing (e.g. Dunn, Baguley & Brunsden, 2014). For example, the extent to which each item measures the same trait; whether the mean and variance (including error variance) are consistent across items.

In examining the results from the Cronbach's alpha further, the results show some variation across the items for means, standard deviations and variance (see Table 1.5). It is likely that this variation is due to varied responses from participants as suggested above. Consequently, future analyses should consider recalculating internal consistency using an alternative method, for example, omega which is more robust when responses are invariant (McDonald, 1999).

Table 1.1. Internal validity of geography and maths classroom cognitive measures and maths non-cognitive measures across the five assessment points for the UK sample, demonstrated by Cronbach's alpha

Measures	Time 1 Cronbach's alpha (N)	Time 2 Cronbach's alpha (N)	Time 3 Cronbach's alpha (N)	Time 4 Cronbach's alpha (N)	Time 5 Cronbach's alpha (N)
Cognitive measures					
Maths performance (MPVT)	.963 (389)	.921 (359)	.922 (347)	.931 (342)	.930 (358)
Geography performance (GPVT)	.922 (389)	.839 (357)	.883 (347)	.886 (342)	.892 (358)
Number line	.808 (303)	.871 (295)	.781 (290)	.881 (291)	.841 (283)
Cognitive ability scores (SAS) *	.905 (141)	-	-	-	-
Non-cognitive measures					
Maths enjoyment	.630 (340)	.734 (350)	.750 (333)	.767 (331)	.776 (345)
Maths self-perceived ability	.619 (341)	.721 (344)	.704 (333)	.739 (328)	.783 (347)
Maths classroom chaos	.766 (267)	.798 (293)	.809 (276)	.805 (288)	.813 (291)
Maths classroom environment total scale	.598 (289)	.736 (308)	.712 (295)	.794 (294)	.814 (301)
Maths student-teacher relation (subscale)	.637 (305)	.814 (322)	.769 (304)	.823 (296)	.835 (311)
Maths Peer competition (subscale)	.520 (316)	.589 (335)	.590 (330)	.684 (329)	.699 (339)
Maths homework total scale	.430 (319)	.584 (343)	.540 (327)	.572 (326)	.553 (328)
Maths homework behaviour (subscale)	.280 (330)	.149 (351)	.160 (340)	.260 (329)	.150 (347)
Maths homework feedback (subscale)	.540 (321)	.722 (346)	.705 (328)	.711 (328)	.754 (330)
Maths environment	.490 (314)	.430 (344)	.564 (328)	.589 (318)	.528 (339)
Maths usefulness	.379 (287)	.554 (334)	.588 (326)	.658 (310)	.619 (340)
Maths anxiety	.837 (284)	.862 (321)	.861 (304)	.896 (308)	.896 (321)
Maths tutoring	.288 (54)	-.067 (53)	-.034 (70)	-.403 (43)	.245 (41)

* data collected from school 1 only. Maths student-teacher relations (8 items) and Maths peer competition (4 items) = subscales of Maths classroom environment total scale (12 items); Maths homework behaviour (2 items) and Maths homework feedback (3 items) = subscales of Maths homework total scale (5 items).

Table 1.2. Internal validity of geography classroom non-cognitive measures across the five assessment points, perceptions of intelligence and socioeconomic status at time 1 and time 4 for the UK sample, demonstrated by Cronbach's alpha

Measures	Time 1 Cronbach's alpha (N)	Time 2 Cronbach's alpha (N)	Time 3 Cronbach's alpha (N)	Time 4 Cronbach's alpha (N)	Time 5 Cronbach's alpha (N)
Geography enjoyment	.423 (252)	.816 (330)	.857 (326)	.858 (320)	.889 (336)
Geography self-perceived ability	.611 (250)	.758 (323)	.826 (218)	.825 (322)	.873 (336)
Geography classroom chaos	.733 (188)	.824 (281)	.846 (247)	.838 (278)	.827 (299)
Geography classroom environment (total scale)	.868 (202)	.842 (282)	.851 (289)	.863 (289)	.868 (305)
Geography student-teacher relation (subscale)	.862 (210)	.866 (293)	.847 (300)	.866 (301)	.875 (311)
Geography peer competition (subscale)	.728 (228)	.731 (305)	.732 (315)	.723 (304)	.752 (327)
Geography homework (total scale)	.646 (213)	.634 (308)	.594 (317)	.608 (309)	.599 (325)
Geography homework behaviour (subscale)	.333 (224)	.205 (316)	.290 (313)	.215 (316)	.240 (332)
Geography homework feedback (subscale)	.777 (214)	.779 (312)	.757 (318)	.755 (313)	.781 (329)
Geography environment	.567 (203)	.678 (301)	.652 (316)	.664 (307)	.750 (330)
Geography usefulness	.643 (186)	.680 (297)	.681 (300)	.702 (301)	.698 (312)
Geography anxiety	.831 (185)	.879 (297)	.870 (299)	.911 (297)	.926 (320)
Geography tutoring	-.181 (12)	-.221 (13)	.717 (20)	.628 (7)	.654 (15)
Theories of intelligence	.805 (327)	-	-	.854 (320)	-
Academic and socioeconomic status	.675 (475)	-	-	.615 (287)	-

Geography student-teacher relations (8 items) and Geography peer competition (4 items) = subscales of Geography classroom environment total scale (12 items); Geography homework behaviour (2 items) and Geography homework feedback (3 items) = subscales of Geography homework total scale (5 items).

Table 1.3. Internal validity of geography and maths classroom cognitive measures and maths non-cognitive measures across the three assessment points for the Russian sample, demonstrated by Cronbach's alpha

	Time 1	Time 2	Time 3
Measures	**Cronbach's alpha (N)**	**Cronbach's alpha (N)**	**Cronbach's alpha (N)**
Cognitive measures			
Maths performance (MPVT)	.880 (229)	.883 (225)	.903 (222)
Geography performance (GPVT)	.828 (229)	.809 (223)	.748 (227)
Number line	.844 (208)	.889 (226)	.812 (205)
Non-cognitive measures			
Maths enjoyment	.588 (214)	.685 (218)	.727 (212)
Maths self-perceived ability	.638 (216)	.702 (218)	.765 (223)
Maths classroom chaos	.717 (171)	.778 (173)	.784 (206)
Maths classroom environment (total scale)	.732 (205)	.798 (217)	.823 (204)
Maths student-teacher relation (subscale)	.759 (212)	.810 (217)	.760 (212)
Maths peer competition (subscale)	.695 (221)	.711 (225)	.695 (215)
Maths homework (total scale)	.504 (221)	.542 (215)	.441 (217)
Maths homework behaviour (subscale)	.430 (225)	.551 (224)	.399 (224)
Maths Homework feedback (subscale)	.491 (223)	.481 (215)	.392 (219)
Maths environment	.166 (212)	.263 (218)	.449 (216)
Maths usefulness	.630 (219)	.738 (214)	.651 (216)
Maths anxiety	.831 (199)	.850 (201)	.876 (210)
Maths tutoring	-.267 (31)	-.629 (44)	.169 (42)

Maths student-teacher relations (8 items) and Maths peer competition (4 items) = subscales of Maths classroom environment total scale (12 items); Maths homework behaviour (2 items) and Maths homework feedback (3 items) = subscales of Maths homework total scale (5 items).

Table 1.4. Internal validity of geography classroom non-cognitive measures across the three assessment points, perceptions of intelligence and socioeconomic status at time 1 for the Russian sample, demonstrated by Cronbach's alpha

	Time 1	Time 2	Time 3
Measures	**Cronbach's alpha (N)**	**Cronbach's alpha (N)**	**Cronbach's alpha (N)**
Geography enjoyment	.608 (210)	.802 (219)	.822 (214)
Geography self-perceived ability	.517 (206)	.761 (215)	.801 (221)
Geography classroom chaos	.732 (193)	.789 (205)	.806 (199)
Geography classroom environment (total scale)	.817 (193)	.830 (207)	.845 (202)
Geography student-teacher relation (subscale)	.801 (197)	.803 (209)	.807 (207)
Geography peer competition (subscale)	.635 (209)	.745 (217)	.769 (214)
Geography homework (total scale)	.548 (207)	.562 (215)	.543 (209)
Geography homework behaviour (subscale)	.335 (212)	.484 (219)	.221 (216)
Geography homework feedback (subscale)	.593 (211)	.630 (216)	.607 (209)
Geography environment	.541 (205)	.509 (211)	.593 (212)
Geography usefulness	.662 (207)	.717 (219)	.735 (212)
Geography anxiety	.839 (184)	.828 (202)	.893 (206)
Geography tutoring	.235 (12)	.710 (17)	.634 (8)
Theories of intelligence	.714 (207)	-	-
Academic and socioeconomic status	.709 (188)	-	-

Geography student-teacher relations (8 items) and Geography peer competition (4 items) = subscales of Geography classroom environment total scale (12 items); Geography homework behaviour (2 items) and Geography homework feedback (3 items) = subscales of Geography homework total scale (5 items).

Table 1.5. Means, standard deviations (SD) and N for individual items of the maths homework total scale at corresponding waves across both Russian and UK samples

	UK sample time 3				Russian sample time 2			
	Mean	**SD**	**Variance**	**N**	**Mean**	**SD**	**Variance**	**N**
1. I complete my homework on time	2.37	0.75	0.54	327	2.39	0.76	0.58	213
2. I do my homework while watching television	2.28	0.92	0.85	327	2.33	0.88	0.77	213
3. My teacher grades my homework	1.52	1.03	1.06	327	2.00	0.87	0.76	213
4. My teacher makes useful comments on my homework	1.68	1.04	1.08	327	1.62	0.96	0.92	213
5. I am given interesting homework	1.11	0.92	0.85	327	1.41	0.94	0.88	213

Descriptive statistics for all assessed variables

Tables 1.6 to 1.8 present student measures and Table 1.9. presents teacher measures.

Table 1.6. Descriptive statistics for maths classroom variables at Time 1, Time 2 and Time 3 for the Russian and UK samples

		Time 1		Time 2		Time 3	
Construct	**Descriptives**	**Russia**	**UK**	**Russia**	**UK**	**Russia**	**UK**
Maths school achievement Time 1 primary, Time 2 year 5 (Russia) end of spring term (UK)	N	219	262	225	281		
	Mean	0.00	0.15	0.00	0.15		
	SD	0.99	0.98	0.99	0.93		
	Skewness	0.20	0.30	0.20	0.26		
	SE Skewness	0.16	0.15	0.16	0.15		
	Kurtosis	-0.43	-0.83	-0.64	-0.20		
	SE Kurtosis	0.33	0.30	0.32	0.29		
	Minimum	-1.94	-2.29	-1.49	-2.20		
	Maximum	2.05	2.15	1.92	2.74		
Maths performance	N	229	290	222	286	220	294
	Mean	0.00	0.11	0.02	0.08	0.00	0.13
	SD	1.00	0.93	0.97	0.96	0.99	0.90
	Skewness	-0.22	-0.44	-0.59	-0.61	-0.48	-0.54
	SE Skewness	0.16	0.14	0.16	0.14	0.16	0.14
	Kurtosis	-0.52	-0.26	-0.24	-0.21	-0.48	-0.21
	SE Kurtosis	0.32	0.29	0.33	0.29	0.33	0.28
	Minimum	-2.85	-2.58	-2.91	-2.55	-2.91	-2.82
	Maximum	2.08	1.71	1.61	1.62	1.78	1.55
Number line	N	226	288	220	282	223	290
	Mean	-0.09	-0.04	-0.04	-0.06	0.10	-0.12
	SD	0.92	0.85	0.92	0.81	0.96	0.81
	Skewness	-0.31	-0.17	-0.56	0.41	-0.04	-0.15
	SE Skewness	0.16	0.14	0.16	0.15	0.16	0.14
	Kurtosis	0.93	0.28	-0.06	1.22	0.50	0.49
	SE Kurtosis	0.32	0.29	0.33	0.29	0.32	0.29
	Minimum	-2.73	-2.93	-2.64	-2.28	-2.73	-2.50
	Maximum	2.97	2.21	1.82	3.00	2.91	2.61
Maths self-perceived ability	N	219	285	216	276	221	287
	Mean	0.02	0.04	0.02	0.06	0.01	0.06
	SD	0.97	0.96	0.98	0.90	0.99	0.89
	Skewness	-0.11	-0.52	-0.11	-0.65	0.10	-0.41
	SE Skewness	0.16	0.14	0.17	0.15	0.16	0.14
	Kurtosis	-0.56	-0.37	-0.27	0.26	-0.49	-0.30
	SE Kurtosis	0.33	0.29	0.33	0.29	0.33	0.29
	Minimum	-2.79	-2.85	-2.91	-2.85	-2.43	-2.55
	Maximum	1.96	1.59	1.87	1.52	2.17	1.42
Maths enjoyment	N	210	284	219	278	212	287
	Mean	0.08	0.03	0.00	0.00	0.00	0.00
	SD	0.88	0.97	1.00	0.98	1.00	0.96
	Skewness	-0.08	-0.19	-0.21	-0.49	-0.30	-0.42
	SE Skewness	0.17	0.14	0.16	0.15	0.17	0.14
	Kurtosis	-0.30	-0.45	-0.03	0.49	0.44	0.39
	SE Kurtosis	0.33	0.29	0.33	0.29	0.33	0.29
	Minimum	-2.75	-2.80	-2.65	-2.88	-2.89	-2.93
	Maximum	1.88	1.97	1.87	1.77	1.97	1.81

All variables are corrected for age and outliers removed, students with special educational needs were excluded from these analyses **SD** = standard deviation; **SE** = standard error; Maths achievement only available at time 1 and 2.

Table 1.6. Continued. Descriptive statistics for maths classroom variables at Time 1, Time 2 and Time 3 for the Russian and UK samples

		Time 1		Time 2		Time 3	
Construct	**Descriptives**	**Russia**	**UK**	**Russia**	**UK**	**Russia**	**UK**
Maths classroom environment	*N*	223	290	222	283	218	290
	Mean	0.03	0.09	0.02	0.06	0.09	0.07
	SD	0.94	0.93	0.97	0.99	0.85	0.94
	Skewness	-0.47	-0.22	-0.37	-0.1	-0.52	-0.2
	SE Skewness	0.16	0.14	0.16	0.14	0.16	0.14
	Kurtosis	0.07	-0.13	-0.15	0.17	0.15	-0.11
	SE Kurtosis	0.32	0.29	0.33	0.29	0.33	0.29
	Minimum	-2.98	-2.89	-2.74	-2.73	-2.47	-2.88
	Maximum	2.03	2.48	2.19	2.74	1.84	2.37
Maths student-teacher relations	*N*	223	291	221	283	218	291
	Mean	0.03	0.07	0.03	0.06	0.08	0.06
	SD	0.95	0.97	0.96	0.97	0.87	0.98
	Skewness	-0.54	-0.33	-0.35	-0.15	-0.29	-0.31
	SE Skewness	0.16	0.14	0.16	0.14	0.16	0.14
	Kurtosis	-0.08	-0.21	-0.25	-0.51	-0.4	-0.4
	SE Kurtosis	0.32	0.28	0.33	0.29	0.33	0.28
	Minimum	-2.86	-2.81	-2.47	-2.3	-2.46	-2.75
	Maximum	1.76	2.22	2.22	2.28	1.77	2.04
Maths peer competition	*N*	223	291	218	283	223	291
	Mean	0.02	0.04	0.07	0.02	0	0.03
	SD	0.96	0.99	0.86	1.01	1	0.95
	Skewness	-0.78	-0.12	-0.93	-0.05	-1.07	-0.1
	SE Skewness	0.16	0.14	0.16	0.14	0.16	0.14
	Kurtosis	-0.17	-0.53	0.21	-0.68	0.5	-0.38
	SE Kurtosis	0.32	0.28	0.33	0.29	0.32	0.28
	Minimum	-2.94	-2.83	-2.67	-2.57	-2.79	-2.35
	Maximum	1.23	2.17	1.11	2.04	0.91	1.88
Maths classroom chaos	*N*	227	290	224	283	227	290
	Mean	0.00	0.05	0.00	0.03	0.00	0.12
	SD	1.00	0.99	1.00	1.01	1.00	0.96
	Skewness	0.02	-0.54	0.10	-0.49	0.04	-1.04
	SE Skewness	0.16	0.14	0.16	0.14	0.16	0.14
	Kurtosis	-0.61	-0.41	-0.87	-0.69	-0.88	0.51
	SE Kurtosis	0.32	0.29	0.32	0.29	0.32	0.29
	Minimum	-2.17	-2.69	-1.93	-2.68	-2.22	-2.91
	Maximum	2.28	1.55	2.39	1.53	2.04	1.41
Maths homework behaviour	*N*	227	289	223	282	224	290
	Mean	-0.03	0.00	-0.07	0.01	0.08	-0.09
	SD	1.05	0.95	1.04	0.95	1.01	0.98
	Skewness	0.25	-0.01	0.26	-0.02	0.10	0.18
	SE Skewness	0.16	0.14	0.16	0.15	0.16	0.14
	Kurtosis	-1.09	-1.11	-1.05	-1.03	-1.08	-1.25
	SE Kurtosis	0.32	0.29	0.32	0.29	0.32	0.29
	Minimum	-1.25	-1.27	-1.30	-1.27	-1.22	-1.25
	Maximum	2.38	2.11	2.28	2.02	2.26	1.99

All variables are corrected for age and outliers removed, students with special educational needs were excluded from these analyses **SD** = standard deviation; **SE** = standard error

Table 1.6. Continued. Descriptive statistics for maths classroom variables at Time 1, Time 2 and Time 3 for the Russian and UK samples

		Time 1		**Time 2**		**Time 3**	
Construct	**Descriptives**	**Russia**	**UK**	**Russia**	**UK**	**Russia**	**UK**
Maths homework feedback	*N*	227	288	220	281	222	288
	Mean	0	0.02	0	0.05	0.01	0.04
	SD	1	1.01	1	0.97	1	0.98
	Skewness	-0.18	0.06	0.16	0.03	0.04	0.24
	SE Skewness	0.16	0.14	0.16	0.15	0.16	0.14
	Kurtosis	-0.19	-0.81	-0.36	-0.69	0.14	-0.64
	SE Kurtosis	0.32	0.29	0.33	0.29	0.33	0.29
	Minimum	-2.87	-2.08	-2.52	-1.93	-2.8	-1.76
	Maximum	1.95	1.9	2	2.01	2.35	1.95
Maths homework total scale	*N*	226	289	221	282	222	288
	Mean	0.02	0.03	0.03	0.06	0.04	0.06
	SD	0.96	0.99	0.97	0.96	0.95	0.96
	Skewness	-0.32	-0.20	-0.38	-0.25	-0.23	0.04
	SE Skewness	0.16	0.14	0.16	0.15	0.16	0.14
	Kurtosis	-0.43	-0.28	0.37	-0.09	0.02	-0.16
	SE Kurtosis	0.32	0.29	0.33	0.29	0.33	0.29
	Minimum	-2.53	-2.75	-2.97	-2.80	-2.65	-2.52
	Maximum	1.96	2.00	2.05	2.21	2.22	2.17
Maths environment	*N*	219	288	218	279	218	288
	Mean	0.00	0.07	0.00	0.07	0.01	0.07
	SD	1.00	0.99	1.00	0.99	0.99	0.96
	Skewness	-0.21	-0.27	-0.09	-0.33	-0.01	-0.42
	SE Skewness	0.16	0.14	0.16	0.15	0.16	0.14
	Kurtosis	-0.44	-0.71	-0.35	-0.57	-0.45	-0.32
	SE Kurtosis	0.33	0.29	0.33	0.29	0.33	0.29
	Minimum	-2.07	-2.18	-2.06	-2.17	-1.96	-2.30
	Maximum	2.28	1.89	2.26	1.67	2.18	1.71
Maths usefulness	*N*	220	287	214	280	221	288
	Mean	-0.05	0.04	-0.06	0.03	-0.03	0.05
	SD	0.91	1.02	0.89	0.99	0.95	0.93
	Skewness	0.56	-0.33	0.64	-0.22	0.62	-0.28
	SE Skewness	0.16	0.14	0.17	0.15	0.16	0.14
	Kurtosis	0.08	-0.27	0.45	-0.21	0.13	-0.06
	SE Kurtosis	0.33	0.29	0.33	0.29	0.33	0.29
	Minimum	-1.32	-2.75	-1.42	-2.84	-1.50	-2.69
	Maximum	3.00	1.78	2.69	1.70	2.98	1.61
Maths anxiety	*N*	220	288	211	282	219	289
	Mean	-0.01	-0.02	-0.02	-0.02	-0.02	0.00
	SD	0.98	1.00	0.97	1.01	0.97	0.99
	Skewness	0.25	0.45	0.17	0.51	0.33	0.39
	SE Skewness	0.16	0.14	0.17	0.15	0.16	0.14
	Kurtosis	-0.75	-0.54	-0.94	-0.04	-0.71	-0.55
	SE Kurtosis	0.33	0.29	0.33	0.29	0.33	0.29
	Minimum	-1.63	-1.65	-1.64	-1.71	-1.47	-1.42
	Maximum	2.68	2.85	2.34	2.97	2.83	2.92

All variables are corrected for age and outliers removed, students with special educational needs were excluded from these analyses **SD** = standard deviation; **SE** = standard error

Table 1.7. Descriptive statistics for geography classroom variables at Time 1, Time 2 and Time 3 for the Russian and UK samples

		Time 1		Time 2		Time 3	
Construct	**Descriptives**	**Russia**	**UK**	**Russia**	**UK**	**Russia**	**UK**
Geography school achievement Time 1 primary, Time 2 year 5	*N*	220		225			
	Mean	0.01		0.01			
	SD	0.99		0.99			
	Skewness	-0.26		-0.18			
	SE Skewness	0.16		0.16			
	Kurtosis	-0.83		-0.90			
	SE Kurtosis	0.33		0.32			
	Minimum	-1.86		-2.04			
	Maximum	1.64		1.32			
Geography performance	*N*	227	288	220	286	224	293
	Mean	0.27	-0.23	0.09	-0.14	0.17	-0.19
	SD	0.82	0.96	0.89	1.02	0.77	1.00
	Skewness	-0.40	-0.29	-0.66	0.04	-0.77	-0.05
	SE Skewness	0.16	0.14	0.16	0.14	0.16	0.14
	Kurtosis	0.97	-0.16	0.30	-0.24	0.83	-0.47
	SE Kurtosis	0.32	0.29	0.33	0.29	0.32	0.28
	Minimum	-2.40	-3.00	-2.30	-2.88	-2.62	-2.61
	Maximum	2.31	2.26	2.15	2.36	1.86	2.21
Geography self-perceptions of ability	*N*	211	266	214	270	222	285
	Mean	0.02	0.03	0.00	-0.01	0.02	0.01
	SD	0.97	1.00	1.00	0.99	0.98	0.97
	Skewness	0.17	-0.50	-0.37	-0.58	0.07	-0.61
	SE Skewness	0.17	0.15	0.17	0.15	0.16	0.14
	Kurtosis	-0.24	0.28	0.28	0.05	0.23	0.05
	SE Kurtosis	0.33	0.30	0.33	0.30	0.33	0.29
	Minimum	-2.00	-2.82	-2.67	-2.53	-2.99	-2.33
	Maximum	2.33	1.73	1.91	1.54	2.16	1.42
Geography enjoyment	*N*	211	272	218	275	215	286
	Mean	0.00	-0.01	0.00	0.00	0.00	0.01
	SD	1.00	1.01	1.00	0.99	1.00	0.98
	Skewness	-0.04	-0.22	-0.40	-0.29	-0.38	-0.28
	SE Skewness	0.17	0.15	0.16	0.15	0.17	0.14
	Kurtosis	0.10	-0.59	-0.09	-0.50	0.13	-0.56
	SE Kurtosis	0.33	0.29	0.33	0.29	0.33	0.29
	Minimum	-2.78	-2.19	-2.31	-2.05	-2.51	-1.85
	Maximum	2.02	1.64	1.65	1.65	1.77	1.56
Geography classroom environment	*N*	214	262	219	270	220	287
	Mean	0.00	0.06	0.02	0.06	0.06	0.04
	SD	1.00	0.97	0.98	0.95	0.91	0.97
	Skewness	-0.42	-0.26	-0.11	0.01	-0.29	0.03
	SE Skewness	0.17	0.15	0.16	0.15	0.16	0.14
	Kurtosis	-0.15	0.08	-0.60	-0.15	-0.62	-0.21
	SE Kurtosis	0.33	0.30	0.33	0.30	0.33	0.29
	Minimum	-2.87	-2.95	-2.59	-2.69	-2.20	-2.47
	Maximum	1.99	2.19	2.05	2.35	1.87	2.55

All variables are corrected for age and outliers removed, students with special educational needs were excluded from these analyses **SD** = standard deviation; **SE** = standard error; Geography school achievement only available for Russian sample

Table 1.7. Continued. Descriptive statistics for geography classroom variables at Time 1, Time 2 and Time 3 for the Russian and UK samples

		Time 1		Time 2		Time 3	
Construct	**Descriptives**	**Russia**	**UK**	**Russia**	**UK**	**Russia**	**UK**
Geography student-teacher relations	*N*	214	262	219	271	223	287
	Mean	0.00	0.04	0.01	0.07	0.01	0.04
	SD	1.00	0.97	0.98	0.98	1.00	0.99
	Skewness	-0.27	-0.21	-0.29	-0.21	-0.48	-0.06
	SE Skewness	0.17	0.15	0.16	0.15	0.16	0.14
	Kurtosis	-0.69	-0.48	-0.69	-0.41	0.00	-0.56
	SE Kurtosis	0.33	0.30	0.33	0.29	0.32	0.29
	Minimum	-2.55	-2.70	-2.39	-2.81	-3.00	-2.34
	Maximum	1.78	1.83	1.99	1.99	1.61	2.12
Geography peer competition	*N*	214	262	219	271	221	286
	Mean	0.00	0.05	0.00	0.00	0.00	0.03
	SD	1.00	1.01	1.00	1.00	1.00	0.99
	Skewness	-0.38	0.07	-0.21	0.05	-0.38	0.22
	SE Skewness	0.17	0.15	0.16	0.15	0.16	0.14
	Kurtosis	-0.11	-0.58	-0.80	-0.53	-0.53	-0.36
	SE Kurtosis	0.33	0.30	0.33	0.29	0.33	0.29
	Minimum	-2.94	-2.11	-2.55	-2.27	-2.53	-2.02
	Maximum	1.77	2.25	1.60	2.21	1.66	2.37
Geography classroom chaos	*N*	218	260	218	268	220	286
	Mean	0.00	0.01	-0.01	0.06	0.00	0.07
	SD	1.00	1.01	1.00	1.00	1.00	0.98
	Skewness	0.06	-0.55	0.05	-0.29	-0.20	-0.65
	SE Skewness	0.16	0.15	0.16	0.15	0.16	0.14
	Kurtosis	-0.89	-0.41	-0.83	-0.98	-0.84	-0.31
	SE Kurtosis	0.33	0.30	0.33	0.30	0.33	0.29
	Minimum	-2.22	-2.63	-2.21	-2.26	-2.32	-2.62
	Maximum	1.99	1.52	2.15	1.60	1.89	1.51
Geography homework behaviour	*N*	216	261	218	268	215	285
	Mean	-0.01	-0.06	-0.02	0.00	0.00	-0.07
	SD	1.01	0.97	1.03	0.96	0.98	0.98
	Skewness	0.31	0.23	0.13	0.00	0.16	0.18
	SE Skewness	0.17	0.15	0.16	0.15	0.17	0.14
	Kurtosis	-1.05	-1.20	-1.34	-1.22	-1.05	-1.29
	SE Kurtosis	0.33	0.30	0.33	0.30	0.33	0.29
	Minimum	-1.18	-1.25	-1.24	-1.22	-1.21	-1.19
	Maximum	2.32	2.08	1.87	2.12	2.25	1.98
Geography homework feedback	*N*	216	260	218	266	214	284
	Mean	0.00	0.01	0.00	0.04	0.01	-0.03
	SD	1.00	0.99	1.00	0.98	1.00	0.99
	Skewness	-0.04	0.16	-0.03	0.15	-0.20	0.37
	SE Skewness	0.17	0.15	0.16	0.15	0.17	0.14
	Kurtosis	-0.59	-0.81	-0.79	-0.74	-0.22	-0.54
	SE Kurtosis	0.33	0.30	0.33	0.30	0.33	0.29
	Minimum	-2.26	-1.72	-2.40	-1.76	-2.49	-1.74
	Maximum	1.85	1.76	1.74	2.02	1.83	2.09

All variables are corrected for age and outliers removed, students with special educational needs were excluded from these analyses **SD** = standard deviation; **SE** = standard error

Table 1.7. Continued. Descriptive statistics for geography classroom variables at Time 1, Time 2 and Time 3 for the Russian and UK samples

		Time 1		Time 2		Time 3	
Construct	**Descriptives**	**Russia**	**UK**	**Russia**	**UK**	**Russia**	**UK**
Geography Homework total scale	*N*	216	260	218	266	213	285
	Mean	0.00	0.04	0.00	0.05	0.02	0.01
	SD	1.00	1.00	1.00	0.99	0.97	0.98
	Skewness	-0.17	-0.07	-0.11	-0.16	-0.32	0.04
	SE Skewness	0.17	0.15	0.16	0.15	0.17	0.14
	Kurtosis	-0.50	-0.42	-0.49	-0.17	-0.04	-0.18
	SE Kurtosis	0.33	0.30	0.33	0.30	0.33	0.29
	Minimum	-2.68	-2.66	-2.59	-2.89	-2.67	-2.76
	Maximum	1.97	1.89	1.99	2.14	2.00	2.26
Geography environment	*N*	207	252	210	262	216	280
	Mean	0.00	-0.04	0.01	0.03	0.00	0.02
	SD	1.00	0.95	1.00	0.97	1.00	1.00
	Skewness	-0.07	-0.11	0.17	-0.06	-0.08	-0.02
	SE Skewness	0.17	0.15	0.17	0.15	0.17	0.15
	Kurtosis	-0.70	-0.69	-0.51	-0.84	-0.69	-0.82
	SE Kurtosis	0.34	0.31	0.33	0.30	0.33	0.29
	Minimum	-2.25	-2.16	-1.69	-1.94	-1.89	-1.68
	Maximum	2.13	1.77	2.34	1.97	2.18	1.95
Geography usefulness	*N*	213	307	220	313	216	329
	Mean	0.00	0.03	0.02	0.03	0.01	0.00
	SD	1.00	0.95	0.98	0.96	1.00	1.00
	Skewness	-0.84	-0.21	-0.40	-0.24	-0.68	-0.23
	SE Skewness	0.17	0.14	0.16	0.14	0.17	0.13
	Kurtosis	1.22	0.39	0.38	0.29	0.69	0.47
	SE Kurtosis	0.33	0.28	0.33	0.27	0.33	0.27
	Minimum	-2.90	-3.00	-2.99	-3.00	-2.77	-2.85
	Maximum	1.76	1.90	1.90	1.86	1.71	1.89
Geography anxiety	*N*	207	266	217	268	221	282
	Mean	-0.04	-0.03	-0.04	-0.04	-0.03	-0.04
	SD	0.93	0.98	0.94	0.95	0.94	0.97
	Skewness	0.57	0.88	0.64	0.73	0.61	1.00
	SE Skewness	0.17	0.15	0.17	0.15	0.16	0.15
	Kurtosis	-0.68	0.12	-0.20	-0.15	-0.32	0.60
	SE Kurtosis	0.34	0.30	0.33	0.30	0.33	0.29
	Minimum	-1.35	-1.33	-1.27	-1.19	-1.28	-1.14
	Maximum	2.36	2.98	2.71	2.74	2.82	2.97

All variables are corrected for age and outliers removed, students with special educational needs were excluded from these analyses **SD** = standard deviation; **SE** = standard error

Table 1.8. Descriptive statistics for perceptions of intelligence, socioeconomic status and cognitive ability at Time 1 for the Russian and UK samples

Country	Descriptives	TOI	SES mean score	Self-perceptions school respect	Self-perceptions school grades	Self-perceptions family occupation	Self-perceptions family education	Cognitive ability
Russian sample	*N*	220	217	204	209	205	205	
	Mean	0.00	0.00	0.00	0.00	0.00	0.00	
	SD	1.00	1.00	1.00	1.00	1.00	1.00	
	Skewness	0.17	0.10	-0.40	-0.31	-0.31	-0.12	
	SE Skewness	0.16	0.17	0.17	0.17	0.17	0.17	
	Kurtosis	-0.30	-0.19	0.19	-0.50	-0.26	-0.47	
	SE Kurtosis	0.33	0.33	0.34	0.33	0.34	0.34	
	Minimum	-2.18	-2.51	-2.81	-2.84	-2.88	-2.55	
	Maximum	2.30	2.10	1.70	1.63	1.65	1.68	
UK sample	*N*	273	271	264	261	253	256	139
	Mean	0.01	0.05	-0.03	0.07	0.01	0.04	0.08
	SD	0.99	1.02	1.00	0.92	0.93	0.91	0.93
	Skewness	-0.23	-0.28	-0.38	-0.28	-0.21	-0.68	-0.22
	SE Skewness	0.15	0.15	0.15	0.15	0.15	0.15	0.21
	Kurtosis	-0.86	-0.26	-0.45	-0.44	0.07	-0.12	-0.44
	SE Kurtosis	0.29	0.29	0.30	0.30	0.31	0.30	0.41
	Minimum	-2.42	-2.98	-2.48	-2.51	-2.50	-2.87	-1.91
	Maximum	1.74	2.32	1.78	1.79	1.78	1.19	2.44

All variables are corrected for age and outliers removed, students with special educational needs were excluded from these analyses. **SD** = standard deviation; **SE** = standard error **TOI** = Theories of intelligence; **SES** = academic and socioeconomic status (composite of self-perceptions of school respect/grades and family occupation/education); cognitive ability only available for UK sample

Table 1.9. Descriptive statistics for teacher characteristics of primary and current subject teachers in the Russian sample

Descriptives	Teacher age at time of testing	How many years have you been teaching since your graduation?	Teacher self efficacy in student engagement mean score	Teacher self efficacy in instructional strategies mean score	Teacher self efficacy in classroom management mean score	Emotional ability mean score
N	14	17	17	17	17	17
Mean	49.93	25.00	6.53	7.49	6.98	5.27
SD	7.87	8.69	1.29	0.78	1.14	0.30
Skewness	-0.18	-0.03	-0.21	-0.19	-0.36	0.01
SE Skewness	0.60	0.55	0.55	0.55	0.55	0.55
Kurtosis	-0.21	-1.23	-0.26	0.01	-0.42	-0.74
SE Kurtosis	1.15	1.06	1.06	1.06	1.06	1.06
Minimum	35	12	4	6	5	4.77
Maximum	63	40	9	9	9	5.80

Raw variables assessed. **SD** = standard deviation; **SE** = standard error

Appendix 2

Factor analyses of classroom environment measure

The measure used to assess perceptions of classroom environment includes questions that relate to different aspects of the classroom environment, for example, teacher-student relations and peer-peer relations. The different questions and low correlations between this and other measures suggest that there may be more than one subscale to the measure. To assess this, factor analysis was conducted on data collected in the UK sample at wave 2 when students had been in the class for 1 term. These findings replicated in the Russian sample, which are not reported here but available from the author on request.

Participants

The sample consisted of 389 11 to 12 year old UK secondary school students (58% males) from two urban mixed ability schools. The students were in Year 7, the first year of their secondary education. They had been in their new class groups at their new schools for one term. The original sample included 70 students with special educational needs that were excluded from these analyses. Variation in N was seen across measures due to absenteeism on days of data collection and data missing at random from students.

Subjective measures of classroom environment

Maths Classroom Environment uses 12 items from a 19 item measure taken from 'Your School' questionnaires used in the TEDS 16 year study. Students are asked to think about their maths classroom environment and teacher in the past year (in this study, since the beginning of term) and rate which statements are true for their classroom. Classroom items include, *some pupils try to be the first ones finished*. Teacher items include, *the teacher shows an interest in every student's learning*. A 4-point scale is used ranging from '0 = never' to '3 = every lesson'. This study uses 12 of the original 19 items, to avoid items inappropriate for this stage of education and prevent overlap with other measures. The original measure is adapted from two questionnaires: Student classroom environment, 9 items adapted from the full 12-item measure (Midgley, Eccles & Feldlaufer, 1991), 10 items from PISA – classroom environment. Chronbach's alpha at wave 2 = .736 (N=308).

Results

All variables were regressed on age to control for any potential age effects and univariate outliers were removed. The mean, standard deviation and distribution are shown in Table 2.1 below. Figure 2.1 confirms that the measure assessing classroom environment at wave 2 is normally distributed.

Table 2.1. Mean, standard deviation and measures of distribution for maths classroom environment at wave 2.

	Maths classroom environment wave 2
N	290
Mean	0.09
Std. Error of Mean	0.05
Median	0.11
Std. Deviation	0.93
Variance	0.86
Skewness	-0.22
Std. Error of Skewness	0.14
Kurtosis	-0.13
Std. Error of Kurtosis	0.29
Range	5.37
Minimum	-2.89
Maximum	2.48

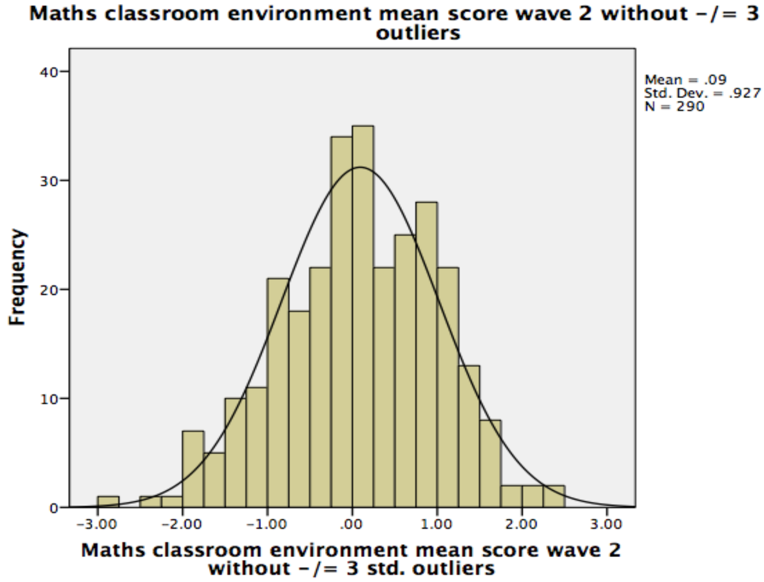

Figure 2.1. Histogram showing the distribution of mean scores for maths classroom environment at wave 2.

A principle component analyses (PCA) was conducted on the 12 items with oblimin rotation (oblique). The Kaiser- Meyer – Olkin measure verified the sampling adequacy for the analysis, KMO = .75 as good, and all KMO values were > .51, which is above the acceptable limit of .5 (Field, 2009). Bartlett's test of sphericity χ^2 (66, 258) = 753.400, p < .001, indicated that correlations between items were sufficiently large for PCA. An initial analysis was run to obtain eigenvalues for each component in the data. Four components had eigenvalues over Kiaser's criterion of 1 and in combination explained 63.29% of the variance. However, three of the components consisted of items loading on two of the components, as shown in Table 2.2 below. Two of the items: *'4. Some pupils tried to be the first ones finished'* and '*3. Some pupils try to be the first ones to answer questions the teacher asks'*, load positively on one component and negatively on another. This suggests that the two scenarios may be viewed positively by some students and unfavourably by other students. All items were retained. Further PCA was conducted and two components were retained in the final analysis that explained 44.12% of the variance. Table 2.3 below shows the factor loadings after the final rotation. Component 1 represents teacher-student relations and class set-up, and component 2 represents peer competition. Items 3 and 4, which load positively on this component, loaded negatively on another factor when four factors were retained.

Reliability analysis was conducted on the two components separately. As with the initial reliability analyses, the whole sample was included. For the eight items that comprise component 1, teacher-student relations, the reliability has increased, Chronbach's alpha = .814 (N=322). For the four items that comprise component 2, peer competition, the reliability has reduced, Chronbach's alpha = .589 (N=335).

Pairwise correlations using the two components showed some increases in associations compared with the whole measure (see Table 2.4). For example, the relationships for student-teacher-class increased with maths enjoyment, classroom chaos and self-perceptions of academic and socioeconomic status. The relationships for peer competition increased with maths problem solving, and school maths achievement. However, the teacher-student-class component did not associate with maths problem solving at all now separated. A decrease was seen between maths anxiety and student-teacher-class, and between maths anxiety and peer competition was non-significant. A decrease was also seen with peer competition and classroom chaos.

In summary, PCA was conducted on the classroom environment measure used to assess teacher-student and student-student relations within the maths classroom at wave 2. With all twelve items retained, the final solution revealed two factors, one related to student-teacher relations and class set up, and the other related to peer competition. The reliability increased for student teacher relations but decreased for peer competition compared with the initial alpha of .736 (N = 308) for all twelve items. It is worth bearing in mind that the measure was initially comprised of nineteen items and several were dropped to avoid overlapping items with other questionnaires. Future analyses may consider further PCA and include these other measures.

Table 2.2. Component matrix from initial PCA showing items loading on more than one component

	Component			
	1	2	3	4
7. The teacher tries to make work interesting in this class	0.785			
10. The teacher shows an interest in every student's learning	0.784			
12. The teacher does a lot to help students	0.729			
11. The teacher gives students an opportunity to express opinions	0.710			
9. The teacher tells us why our work is important	0.652			
8. The teacher likes the work she/he gives us	0.584			
4. Some pupils try to be the first ones finished		0.751	-0.453	
3. Some pupils try to be the first ones to answer questions the teacher asks		0.712	-0.499	
6. When we get reports, we tell each other what we got		0.672	0.524	
5. When work is handed back, we show each other how we did		0.484	0.707	
2. We help each other with our work	0.421			0.670
1. We get to work with each other in small groups	0.488			0.619

Extraction Method: Principal Component Analysis.
a. 4 components extracted.

Table 2.3. Summary of exploratory factor analysis results for maths classroom environment measure ($N = 258$).

Item	Rotated Factor Loadings 1	2
7. The teacher tries to make work interesting in this class	0.785	
10. The teacher shows an interest in every student's learning	0.784	
12. The teacher does a lot to help students	0.729	
11. The teacher gives students an opportunity to express opinions	0.710	
9. The teacher tells us why our work is important	0.652	
8. The teacher likes the work she/he gives us	0.584	
1. We get to work with each other in small groups	0.488	
2. We help each other with our work	0.421	
4. Some pupils try to be the first ones finished		0.751
3. Some pupils try to be the first ones to answer questions the teacher asks		0.712
6. When we get reports, we tell each other what we got		0.672
5. When work is handed back, we show each other how we did		0.484
Eigen values	3.464	1.83
% of Variance	28.866	15.25

Extraction Method: Principal Component Analysis. a. 2 components extracted.

Table 2.4. Bivariate correlations between maths classroom environment, maths classroom teacher-student relations, maths classroom peer competition, maths problem solving, self-perceptions of maths ability (SPA), maths enjoyment, School maths achievement, maths classroom chaos, maths anxiety at wave 2 and self-perceptions of academic and socioeconomic status at wave 1 (ASES), (N).

	1	2	3	4	5	6	7	8	9
1. Maths performance	-								
2. Maths SPA	**.419**** **(284)**	-							
3. Maths enjoyment	**.292**** **(283)**	**.693**** **(282)**	-						
4. Maths classroom environment	**.161**** **(289)**	.102 (284)	**.210**** **(283)**	-					
5. Maths classroom student-teacher-class	0.1 (290)	.089 (285)	**.258**** **(284)**	**.873**** **(290)**	-				
6. Maths classroom peer competition	**.165**** **(290)**	.004 (285)	.018 (284)	**.511**** **(290)**	.056 (291)	-			
7. School maths achievement	**.645**** **(287)**	**.267**** **(282)**	**.147*** **(281)**	.083 (287)	-0.01 (288)	**.185**** **(288)**	-		
8. Maths classroom chaos	.064 (289)	.065 (284)	**.132*** **(283)**	**.247**** **(290)**	**.379**** **(290)**	**-.156**** **(290)**	.032 (287)	-	
9. Maths anxiety	**-.256**** **(287)**	**-.182**** **(282)**	-.060 (281)	**.136*** **(288)**	**.123*** **(288)**	0.1 (288)	**-.263**** **(285)**	**-.141*** **(288)**	-
10. ASES	**.263**** **(265)**	**.311**** **(260)**	**.213**** **(259)**	.114 (266)	**.123*** **(266)**	0.02 (266)	**.125*** **(263)**	**.143*** **(266)**	-.104 (265)

** Correlation is significant at the 0.01 level (2-tailed).* Correlation is significant at the 0.05 level (2-tailed). Bold = significant

Appendix 3

Supplementary materials for Chapter 3

Table 3.1. Levene's tests of equality of variances for maths classroom measures

Construct	Time	F	df1	df2	Sig.
Maths performance	Time 1	1.992	1	423	.159
	Time 2	.100	1	423	.752
	Time 3	5.047	1	423	.025
Number line	Time 1	.389	1	413	.533
	Time 2	6.071	1	413	**.014**
	Time 3	7.416	1	413	**.007**
Maths self-perceived ability	Time 1	.074	1	392	.786
	Time 2	2.327	1	392	.128
	Time 3	3.088	1	392	.080
Maths enjoyment	Time 1	2.386	1	381	.123
	Time 2	.111	1	381	.740
	Time 3	.020	1	381	.888
Maths classroom environment	Time 1	.070	1	412	.791
	Time 2	.394	1	412	.530
	Time 3	2.213	1	412	.138
Maths classroom student-teacher relations	Time 1	.285	1	414	.594
	Time 2	.023	1	414	.880
	Time 3	5.125	1	414	**.024**
Maths classroom peer competition	Time 1	.083	1	412	.774
	Time 2	2.399	1	412	.122
	Time 3	.188	1	412	.665
Maths classroom chaos	Time 1	.127	1	421	.722
	Time 2	.055	1	421	.815
	Time 3	.940	1	421	.333
Maths homework behaviour	Time 1	.724	1	418	.395
	Time 2	2.570	1	418	.110
	Time 3	.543	1	418	.462
Maths homework feedback	Time 1	2.097	1	409	.148
	Time 2	.057	1	409	.812
	Time 3	.254	1	409	.615

Significant results in bold following a Bonferroni multiple testing correction of $p \leq .001$ ($p = .05$ divided by number of measures (k=90) across maths and geography, Time 1, 2 & 3.

Table 3.2. Levene's tests of equality of variances for maths classroom measures

Construct	Time	F	df1	df2	Sig.
Maths homework total scale	Time 1	1.997	1	412	.158
	Time 2	.848	1	412	.358
	Time 3	1.011	1	412	.315
Maths environment	Time 1	.197	1	401	.657
	Time 2	.094	1	401	.759
	Time 3	.194	1	401	.660
Maths usefulness	Time 1	4.903	1	399	**.027**
	Time 2	2.592	1	399	.108
	Time 3	.369	1	399	.544
Maths anxiety	Time 1	.009	1	399	.925
	Time 2	.089	1	399	.766
	Time 3	.495	1	399	.482

Significant results in bold following a Bonferroni multiple testing correction of $p \leq .001$ ($p = .05$ divided by number of measures (k=90) across maths and geography, Time 1, 2 & 3.

Table 3.3. Mauchly's tests of sphericity for the within-participants effect of time for maths classroom measures

Construct	X^2	df	Sig.
Maths performance	.663	2	.718
Number line	.961	2	.618
Maths self-perceived ability	3.887	2	.143
Maths enjoyment	.052	2	.974
Maths classroom environment	1.327	2	.515
Maths classroom student-teacher relations	3.189	2	.203
Maths classroom peer competition	1.976	2	.372
Maths classroom chaos	8.338	2	.015
Maths homework behaviour	17.500	2	**.000**
Maths homework feedback	9.131	2	.010
Maths homework total scale	11.169	2	.004
Maths environment	5.637	2	.060
Maths usefulness	.190	2	.909
Maths anxiety	6.623	2	.036

Significant results in bold following a Bonferroni multiple testing correction of $p \leq .001$ ($p = .05$ divided by number of measures (k=90) across maths and geography, Time 1, 2 & 3.

Table 3.4. Levene's tests of equality of variances for geography classroom measures

Construct		F	df1	df2	Sig.
Geography performance	Time 1	6.090	1	419	**.014**
	Time 2	2.525	1	419	.113
	Time 3	13.437	1	419	**.000**
Geography self perceptions of ability	Time 1	.083	1	367	.774
	Time 2	.486	1	367	.486
	Time 3	.175	1	367	.676
Geography enjoyment	Time 1	.004	1	376	.948
	Time 2	.690	1	376	.407
	Time 3	.085	1	376	.770
Geography classroom environment	Time 1	.116	1	372	.733
	Time 2	.040	1	372	.841
	Time 3	3.430	1	372	.065
Geography classroom student-teacher	Time 1	.186	1	375	.667
	Time 2	.034	1	375	.853
	Time 3	1.101	1	375	.295
Geography classroom peer competition	Time 1	.003	1	374	.958
	Time 2	.201	1	374	.654
	Time 3	.421	1	374	.517
Geography classroom chaos	Time 1	.004	1	369	.949
	Time 2	.048	1	369	.826
	Time 3	.001	1	369	.979

Significant results in bold following a Bonferroni multiple testing correction of $p \leq .001$ ($p = .05$ divided by number of measures (k=90) across maths and geography, Time 1, 2 & 3.

Table 3.5. Levene's tests of equality of variances for geography classroom measures

Construct		F	df1	df2	Sig.
Geography homework behaviour	Time 1	.206	1	364	.650
	Time 2	1.260	1	364	.262
	Time 3	.021	1	364	.156
Geography homework feedback	Time 1	.055	1	361	.814
	Time 2	.020	1	361	.888
	Time 3	.002	1	361	.962
Geography homework total scale	Time 1	.012	1	361	.914
	Time 2	.143	1	361	.705
	Time 3	.417	1	361	.519
Geography environment	Time 1	.152	1	343	.697
	Time 2	1.119	1	343	.291
	Time 3	.094	1	343	.760
Geography usefulness	Time 1	.780	1	350	.378
	Time 2	2.091	1	350	.149
	Time 3	.127	1	350	.722
Geography anxiety	Time 1	1.391	1	365	.239
	Time 2	.796	1	365	.373
	Time 3	.624	1	365	.430

Significant results in bold following a Bonferroni multiple testing correction of $p \leq .001$ ($p = .05$ divided by number of measures (k=90) across maths and geography, Time 1, 2 & 3

Table 3.6. Mauchly's tests of sphericity for the within-participants effect of time for maths classroom measures

Construct	χ^2	df	Sig.
Geography performance	2.447	2	.294
Geography self perceptions of ability	.845	2	.655
Geography enjoyment	5.804	2	.055
Geography classroom environment	9.733	2	.008
Geography classroom student-teacher	5.828	2	.054
Geography classroom peer competition	1.097	2	.578
Geography classroom chaos	8.697	2	.013
Geography homework behaviour	9.125	2	**.010**
Geography homework feedback	5.919	2	.052
Geography homework total scale	8.750	2	**.013**
Geography environment	4.528	2	.104
Geography usefulness	2.207	2	.332
Geography anxiety	5.524	2	.063

Significant results in bold following a Bonferroni multiple testing correction of $p \leq .001$ ($p = .05$ divided by number of measures (k=90) across maths and geography, Time 1, 2 & 3. All measures collected at time 1 only for both countries.

Table 3.7. Levene's tests of equality of variances for perceptions of intelligence and socioeconomic status at time 1

Construct	**F**	**df1**	**df2**	**Sig.**
Theories of intelligence	.449	1	491	.503
Perceptions of academic and socioeconomic status	.878	1	486	.349
Self-perceptions of school respect	.282	1	466	.595
Self-perceptions of school grades	2.465	1	468	.117
Self-perceptions of family SES, occupation	.878	1	456	.349
Self-perceptions of family SES, education	3.505	1	459	.062

Significant results in bold following a Bonferroni multiple testing correction of $p \leq .001$ ($p = .05$ divided by number of measures (k=90) across maths and geography, Time 1, 2 & 3. All measures collected at time 1 only for both countries.

Appendix 4

Supplementary materials for Chapter 4

Table 4.1. Levene's test of equality of variances for school 1 maths classroom measures by classroom at time 1

Construct	F	df1	df2	Sig.
Maths primary achievement	2.036	7	164	**.054**
Maths performance	2.301	7	178	**.029**
Number line	1.497	7	175	.171
Maths self-perceived ability	.821	7	173	.571
Maths enjoyment	.505	7	165	.830
Maths classroom environment	3.081	7	176	**.004**
Maths student-teacher relations	2.689	7	176	**.011**
Maths peer competition	1.986	7	176	.059
Maths classroom chaos	.806	7	176	.584
Maths homework behaviour	.506	7	176	.829
Maths homework feedback	1.401	7	176	.208
Maths homework total scale	.787	7	175	.599
Maths environment	1.462	7	171	.184
Maths usefulness	2.014	7	170	.056
Maths anxiety	.933	7	171	.483

Bold = significant at $p \leq .05$

Table 4.2. Levene's test of equality of variances for school 2 maths classroom measures by classroom at time 1

Construct	F	df1	df2	Sig.
Maths primary achievement	.014	2	32	.986
Maths performance	1.192	2	40	.314
Number line	5.566	2	40	**.007**
Maths self-perceived ability	1.169	2	35	.322
Maths enjoyment	.649	2	34	.529
Maths classroom environment	.474	2	36	.626
Maths student-teacher relations	.281	2	36	.756
Maths peer competition	4.429	2	36	**.019**
Maths classroom chaos	.678	2	40	.513
Maths homework behaviour	3.671	2	40	**.034**
Maths homework feedback	.838	2	40	.440
Maths homework total scale	2.355	2	40	.108
Maths environment	.198	2	37	.821
Maths usefulness	.303	2	39	.740
Maths anxiety	2.124	2	38	.134

Bold = significant at $p \leq .05$

Table 4.3. Levene's test of equality of variances for school 1 geography classroom measures by classroom at time 1

Construct	F	df1	df2	Sig.
Geography primary achievement	1.166	7	165	.325
Geography performance	1.461	7	172	.184
Geography self-perceived ability	1.717	7	161	.108
Geography enjoyment	.711	7	162	.663
Geography classroom environment	3.268	7	166	**.003**
Geography student-teacher relations	2.530	7	166	**.017**
Geography peer competition	1.508	7	166	.168
Geography classroom chaos	1.056	7	168	.394
Geography homework behaviour	1.589	2	39	.217
Geography homework feedback	1.078	7	166	.380
Geography homework total scale	1.119	7	166	.353
Geography environment	.727	7	160	.649
Geography usefulness	1.437	7	164	.194
Geography anxiety	1.341	7	161	.234

Bold = significant at $p \leq .05$

Table 4.4. Levene's test of equality of variances for school 2 geography classroom measures by classroom at time 1

Construct	F	df1	df2	Sig.
Geography primary achievement	3.541	2	32	**.041**
Geography performance	1.176	2	40	.319
Geography self-perceived ability	1.539	2	39	.227
Geography enjoyment	2.941	2	38	.065
Geography classroom environment	5.810	2	37	**.006**
Geography student-teacher relations	4.246	2	37	**.022**
Geography peer competition	.582	2	37	.564
Geography classroom chaos	.759	2	39	.475
Geography homework behaviour	1.589	2	39	.217
Geography homework feedback	.855	2	39	.433
Geography homework total scale	8.817	2	39	**.001**
Geography environment	.872	2	36	.427
Geography usefulness	2.615	2	38	.086
Geography anxiety	3.770	2	35	**.033**

Bold = significant at $p \leq .05$

Table 4.5. Levene's test of equality of variances for school 1 and school 2 perceptions of intelligence and academic and socioeconomic status measures by classroom at time 1

School	Construct	F	df1	df2	Sig.
	Theories of intelligence	1.137	7	170	.343
	SES	1.731	7	169	.105
	Self-perceptions of school respect	.298	7	158	.954
	Self-perceptions of school grades	1.006	7	162	.429
	Self-perceptions of family SES, occupation	1.052	7	157	.397
School 1	Self-perceptions of family SES, education	.631	7	160	.730
	Theories of intelligence	.902	2	39	.414
	SES	2.614	2	37	.087
	Self-perceptions of school respect	4.405	2	35	**.020**
	Self-perceptions of school grades	1.659	2	36	.204
	Self-perceptions of family SES, occupation	.182	2	37	.835
School 2	Self-perceptions of family SES, education	1.552	2	34	.226

Bold = significant at $p \leq .05$; **SES** = Perceptions of academic and socioeconomic status: composite of self-perceptions of school respect/grades and family occupation/education

Table 4.6. Levene's test of equality of variances for perceptions of intelligence, and academic and socioeconomic status measures by maths teacher at time 1

Construct	F	df1	df2	Sig.
Maths primary achievement	1.009	5	201	.414
Maths performance	3.075	5	223	**.011**
Number line	3.642	5	220	**.003**
Maths self-perceived ability	.500	5	213	.776
Maths enjoyment	.413	5	204	.839
Maths classroom environment	1.628	5	217	.154
Maths student-teacher relations	1.900	5	217	.095
Maths peer competition	.729	5	217	.603
Maths classroom chaos	1.889	5	221	.097
Maths homework behaviour	.401	5	221	.848
Maths homework feedback	2.346	5	221	.042
Maths homework total scale	1.381	5	220	.233
Maths environment	1.943	5	213	.088
Maths usefulness	1.856	5	214	.103
Maths anxiety	.590	5	214	.708
Theories of intelligence	.920	5	214	.469
Perceptions of academic and socioeconomic status	1.557	5	211	.174
Self-perceptions of school respect	.752	5	198	.585
Self-perceptions of school grades	1.062	5	203	.383
Self-perceptions of family SES, occupation	1.915	5	199	.093
Self-perceptions of family SES, education	1.148	5	199	.337

Bold = significant at $p \leq .05$

Table 4.7. Levene's test of equality of variances for perceptions of intelligence, and academic and socioeconomic status measures by geography teacher at time 1

Construct	F	df1	df2	Sig.
Geography primary achievement	2.400	4	203	**.051**
Geography performance	.402	4	218	.807
Geography self-perceived ability	2.703	4	206	**.032**
Geography enjoyment	.168	4	206	.954
Geography classroom environment	3.485	4	209	**.009**
Geography student-teacher relations	3.296	4	209	**.012**
Geography peer competition	.466	4	209	.760
Geography classroom chaos	2.147	4	213	.076
Geography homework behaviour	.695	4	211	.596
Geography homework feedback	1.530	4	211	.195
Geography homework total scale	1.344	4	211	.255
Geography environment	.776	4	202	.542
Geography usefulness	1.999	4	208	.096
Geography anxiety	2.389	4	202	.052
Theories of intelligence	.459	4	215	.765
Perceptions of academic and socioeconomic status	2.495	4	212	**.044**
Self-perceptions of school respect	.876	4	199	.479
Self-perceptions of school grades	1.703	4	204	.151
Self-perceptions of family SES, occupation	2.833	4	200	**.026**
Self-perceptions of family SES, education	.656	4	200	.624

Bold = significant at $p \leq .05$

Table 4.8. Levene's test of equality of variances for primary school achievement measures by linguistic group at time 1

Construct	F	df1	df2	Sig.
Maths primary achievement	3.187	2	204	**.043**
Geography primary achievement	4.477	2	205	**.013**

Bold = significant at $p \leq .05$

Appendix 5

Supplementary materials for Chapter 5

Table 5.1.1. Levene's test of equality of variances for school 1 maths classroom measures without controlling for prior achievement at time 2 and time 3

Wave	Construct	F	df1	df2	Sig.
Time 2	Maths Year 5 achievement	.603	7	165	.753
	Maths performance	2.149	7	177	**.041**
	Number line	1.327	7	175	.240
	Maths self-perceived ability	1.378	7	171	.217
	Maths enjoyment	1.490	7	175	.174
	Maths classroom environment	1.181	7	177	.316
	Maths student-teacher relations	.952	7	177	.468
	Maths peer competition	.811	7	175	.580
	Maths classroom chaos	2.375	7	178	**.024**
	Maths homework behaviour	.537	7	177	.806
	Maths homework feedback	2.332	7	175	**.027**
	Maths homework total scale	1.202	7	176	.304
	Maths environment	1.827	7	173	.085
	Maths usefulness	1.012	7	168	.424
	Maths anxiety	1.260	7	165	.273
Time 3	Maths performance	1.738	7	165	.103
	Number line	1.414	7	169	.203
	Maths self-perceived ability	.397	7	169	.903
	Maths enjoyment	1.728	7	162	.106
	Maths classroom environment	1.409	7	168	.205
	Maths student-teacher relations	1.346	7	168	.232
	Maths peer competition	2.229	7	170	**.034**
	Maths classroom chaos	.902	7	172	.507
	Maths homework behaviour	1.749	7	169	.101
	Maths homework feedback	1.045	7	168	.402
	Maths homework total scale	1.007	7	167	.428
	Maths environment	1.154	7	167	.332
	Maths usefulness	.701	7	166	.671
	Maths anxiety	1.002	7	167	.431

Bold = significant at $p \leq .05$

Table 5.1.2. Levene's test of equality of variances for school 2 maths classroom measures without controlling for prior achievement at time 2 and time 3

Wave	Construct	F	df1	df2	Sig.
Time 2	Maths Year 5 achievement	.091	2	27	.913
	Maths performance	1.474	2	33	.244
	Number line	2.443	2	33	.102
	Maths self-perceived ability	2.467	2	33	.100
	Maths enjoyment	1.908	2	32	.165
	Maths classroom environment	3.277	2	33	**.050**
	Maths student-teacher relations	2.822	2	32	.074
	Maths peer competition	2.108	2	31	.138
	Maths classroom chaos	1.404	2	34	.260
	Maths homework behaviour	.791	2	34	.461
	Maths homework feedback	4.295	2	33	**.022**
	Maths homework total scale	2.461	2	33	.101
	Maths environment	.155	2	33	.857
	Maths usefulness	2.688	2	34	.082
	Maths anxiety	5.339	2	34	**.010**
Time 3	Maths performance	2.921	2	35	.067
	Number line	.568	2	34	.572
	Maths self-perceived ability	2.710	2	32	.082
	Maths enjoyment	1.151	2	31	.329
	Maths classroom environment	.872	2	32	.428
	Maths student-teacher relations	.714	2	32	.497
	Maths peer competition	2.849	2	34	.072
	Maths classroom chaos	4.056	2	35	.026
	Maths homework behaviour	1.314	2	35	.282
	Maths homework feedback	5.621	2	35	.008
	Maths homework total scale	4.454	2	35	.019
	Maths environment	.223	2	33	.801
	Maths usefulness	.781	2	35	.466
	Maths anxiety	5.502	2	33	.009

Bold = significant at $p \leq .05$

Table 5.1.3. Levene's test of equality of variances for school 1 geography classroom measures without controlling for prior achievement at time 2 and time 3

Wave	Construct	F	df1	df2	Sig.
Time 2	Geography Year 5 achievement	1.442	7	165	.192
	Geography performance	.741	7	173	.637
	Geography self-perceived ability	1.331	7	169	.238
	Geography enjoyment	.393	7	172	.906
	Geography classroom environment	1.676	7	173	.118
	Geography student-teacher relations	2.555	7	173	**.016**
	Geography peer competition	1.840	7	173	.082
	Geography classroom chaos	1.204	7	172	.303
	Geography homework behaviour	5.101	2	33	.012
	Geography homework feedback	2.519	7	173	**.017**
	Geography homework total scale	1.842	7	173	.082
	Geography environment	.718	7	165	.657
	Geography usefulness	1.668	7	175	.120
	Geography anxiety	.729	7	173	.647
Time 3	Geography performance	0.636	7	170	.726
	Geography self-perceived ability	.456	7	169	.865
	Geography enjoyment	2.020	7	164	.055
	Geography classroom environment	2.378	7	167	.024
	Geography student-teacher relations	.747	7	168	.633
	Geography peer competition	1.627	7	167	.131
	Geography classroom chaos	2.834	7	165	.008
	Geography homework behaviour	.396	7	164	.904
	Geography homework feedback	1.469	7	164	.181
	Geography homework total scale	1.508	7	163	.168
	Geography environment	1.143	7	163	.339
	Geography usefulness	1.289	7	162	.259
	Geography anxiety	0.977	7	167	.450

Bold = significant at $p \leq .05$

Table 5.1.4. Levene's test of equality of variances for school 2 geography classroom measures without controlling for prior achievement at time 2 and time 3

Wave	Construct	F	df1	df2	Sig.
Time 2	Geography Year 5 achievement	.455	2	27	.639
	Geography performance	.150	2	33	.861
	Geography self-perceived ability	1.793	2	33	.182
	Geography enjoyment	.525	2	34	.596
	Geography classroom environment	1.429	2	34	.254
	Geography S-T relations	2.517	2	34	.096
	Geography peer competition	.153	2	34	.859
	Geography classroom chaos	3.168	2	34	.055
	Geography homework behaviour	5.101	2	33	.012
	Geography homework feedback	.898	2	33	.417
	Geography homework total scale	2.697	2	33	.082
	Geography environment	1.708	2	33	.197
	Geography usefulness	.057	2	33	.945
	Geography anxiety	.520	2	32	.599
Time 3	Geography performance	1.440	2	34	.251
	Geography self-perceived ability	2.766	2	34	.077
	Geography enjoyment	1.629	2	32	.212
	Geography classroom environment	.255	2	34	.776
	Geography S-T relations	.386	2	35	.682
	Geography peer competition	1.298	2	34	.286
	Geography classroom chaos	2.208	2	35	.125
	Geography homework behaviour	2.017	2	32	.150
	Geography homework feedback	.581	2	31	.565
	Geography homework total scale	1.365	2	31	.270
	Geography environment	1.333	2	34	.277
	Geography usefulness	.330	2	34	.721
	Geography anxiety	.738	2	34	.486

Bold = significant at $p \leq .05$. **S-T** = student-teacher

Table 5.1.5. Levene's test of equality of variances for maths classroom measures by maths teacher without controlling for prior achievement at time 2 and time 3

Wave	Construct	F	df1	df2	Sig.
Time 2	Maths Year 5 achievement	.813	5	197	.542
	Maths performance	1.668	5	215	.144
	Number line	.233	5	213	.948
	Maths self-perceived ability	.563	5	209	.729
	Maths enjoyment	.142	5	212	.982
	Maths classroom environment	2.179	5	215	.058
	Maths student-teacher relations	1.667	5	214	.144
	Maths peer competition	.842	5	211	.521
	Maths classroom chaos	2.874	5	217	**.016**
	Maths homework behaviour	.618	5	216	.686
	Maths homework feedback	.312	5	213	.906
	Maths homework total scale	.249	5	214	.940
	Maths environment	1.876	5	211	.100
	Maths usefulness	1.058	5	207	.385
	Maths anxiety	1.325	5	204	.255
Time 3	Maths performance	.867	5	205	.504
	Number line	.718	5	208	.611
	Maths self-perceived ability	.720	5	206	.609
	Maths enjoyment	1.175	5	198	.323
	Maths classroom environment	3.564	5	205	**.004**
	Maths student-teacher relations	3.564	5	205	**.004**
	Maths peer competition	2.837	5	209	**.017**
	Maths classroom chaos	1.136	5	212	.342
	Maths homework behaviour	2.417	5	209	.037
	Maths homework feedback	1.463	5	208	.203
	Maths homework total scale	1.626	5	207	.155
	Maths environment	.280	5	205	.924
	Maths usefulness	.814	5	206	.541
	Maths anxiety	1.736	5	205	.128

Bold = significant at $p \leq .05$

Table 5.1.6. Levene's test of equality of variances for geography classroom measures by geography teacher without controlling for prior achievement at time 2 and time 3

Wave	Construct	F	df1	df2	Sig.
Time 2	Geography Year 5 achievement	.583	4	198	.675
	Geography performance	1.068	4	212	.373
	Geography self-perceived ability	2.728	4	208	**.030**
	Geography enjoyment	.860	4	212	.489
	Geography classroom environment	.872	4	213	.482
	Geography S-T relations	1.162	4	213	.329
	Geography peer competition	.526	4	213	.716
	Geography classroom chaos	1.353	4	212	.251
	Geography homework behaviour	.685	4	212	.603
	Geography homework feedback	1.899	4	212	.112
	Geography homework total scale	1.280	4	212	.279
	Geography environment	.341	4	204	.850
	Geography usefulness	1.776	4	214	.135
	Geography anxiety	.966	4	211	.427
Time 3	Geography performance	1.332	4	210	.259
	Geography self-perceived ability	.739	4	209	.566
	Geography enjoyment	2.648	4	202	**.035**
	Geography classroom environment	2.499	4	207	**.044**
	Geography S-T relations	.918	4	209	.454
	Geography peer competition	2.429	4	207	**.049**
	Geography classroom chaos	3.378	4	206	**.011**
	Geography homework behaviour	.079	4	202	.989
	Geography homework feedback	1.285	4	201	.277
	Geography homework total scale	.699	4	200	.593
	Geography environment	.616	4	203	.651
	Geography usefulness	1.448	4	202	.220
	Geography anxiety	4.090	4	207	**.003**

Bold = significant at $p \leq .05$. **S-T** = student-teacher

Table 5.1.7. Students' (N) classes and their primary, maths and geography teachers showing their teachers' years of experience (exp.)

	C1e	C2e	C3e	C4se	C5se	C6se	C7se	C8se	C9ce	C10ce	C11ce
N	23	9	18	28	25	28	24	31	18	11	14
Primary Teacher	12 exp.	40 exp.	29 exp.	-	27 exp.	33 exp.	22 exp.	14 exp.	34 exp.	16 exp.	Retired
Maths Teacher	40 exp. TM3	40 exp. TM3	14 exp. TM4	36 exp. TM6	36 exp. TM6	28 exp. TM5	36 exp. TM6	36 exp. TM6	32 exp. TM1	- TM2	- TM2
Geography Teacher	16 exp. TG4	16 exp. TG4	- TG5	- TG5	27 exp. TG2	18 exp. TG3	27 exp. TG2	27 exp. TG2	14 exp. TG1	14 exp. TG1	14 exp. TG1

Table 5.2.6. Maths classroom variables at time 2 for school 1 (Russian sample): Means, standard deviation (SD), and N with ANOVA results by classroom, controlling for prior achievement

Construct	C1e	C2e	C3e	C4se	C5se	C6se	C7se	C8se	*p*	η_p^2
Maths Year 5 school achievement	0.40 (0.94) n=19	0.31 (0.97) n=20	0.15 (0.74) n=14	-0.15 (1.05) n=22	0.18 (1.01) n=20	0.40 (0.76) n=28	-0.06 (0.98) n=20	-0.84 (1.18) n=22	**.000**	**.151**
Maths performance	0.11 (1.01) n=19	-0.10 (1.16) n=20	0.13 (1.02) n=14	0.19 (1.04) n=22	-0.63 (1.01) n=20	0.54 (0.73) n=28	-0.07 (1.03) n=22	-0.56 (0.76) n=22	.001	.142
Number line	0.02 (0.90) n=19	0.38 (0.92) n=20	-0.19 (0.90) n=14	-0.52 (1.07) n=22	0.36 (0.75) n=20	-0.11 (0.91) n=27	-0.10 (0.73) n=22	0.88 (0.62) n=21	**.000**	**.190**
Maths self-perceived ability	0.55 (0.88) n=18	0.07 (0.89) n=20	-0.04 (1.01) n=13	0.06 (0.91) n=21	-0.19 (0.84) n=20	0.15 (1.11) n=27	0.37 (0.82) n=22	-0.53 (1.31) n=22	.031	.093
Maths enjoyment	0.56 (1.09) n=17	0.13 (0.62) n=20	-0.17 (1.10) n=13	0.08 (0.82) n=22	-0.01 (0.98) n=20	-0.02 (1.11) n=28	0.04 (1.07) n=22	-0.26 (0.95) n=22	.366	.047
Maths classroom environment	-0.05 (0.75) n=19	0.41 (0.71) n=20	0.38 (1.04) n=14	0.26 (0.89) n=21	-0.21 (1.27) n=20	0.24 (0.86) n=28	0.11 (0.86) n=22	-0.13 (0.96) n=22	.276	.053
Maths classroom Student-teacher relations	0.06 (0.71) n=19	0.43 (0.73) n=20	0.32 (1.10) n=14	0.19 (0.88) n=21	-0.32 (1.21) n=20	0.22 (0.92) n=28	0.12 (0.87) n=22	-0.19 (0.97) n=22	.188	.060
Maths classroom Peer competition	-0.35 (1.06) n=18	-0.01 (0.80) n=20	0.31 (0.75) n=14	0.36 (0.75) n=21	0.35 (1.08) n=20	0.12 (0.81) n=28	-0.07 (1.05) n=22	0.18 (1.08) n=22	.259	.054

Significant results in bold following a Bonferroni multiple testing correction of $p \leq .000$ ($p = .05$ divided by number of measures (k=114) across maths and geography, time 1, 2 & 3)

Table 5.2.6. Continued. Maths classroom variables at time 2 for school 1 (Russian sample): Means, standard deviation (SD), and N with ANOVA results by classroom, controlling for prior achievement

Construct	C1e	C2e	C3e	C4se	C5se	C6se	C7se	C8se	*p*	η_p^2
Maths classroom chaos	-0.21 (0.82) n=19	0.35 (0.66) n=20	0.53 (1.30) n=14	-0.61 (1.06) n=22	-0.56 (0.89) n=20	0.17 (0.79) n=28	0.60 (0.90) n=22	-0.27 (0.93) n=22	**.000**	**.194**
Maths homework behaviour	-0.16 (0.98) n=19	0.27 (0.99) n=20	0.27 (1.01) n=13	-0.20 (0.87) n=22	0.07 (1.06) n=20	-0.16 (1.06) n=28	-0.26 (0.96) n=22	0.08 (1.24) n=22	.573	.035
Maths homework feedback	-0.58 (0.94) n=19	0.51 (0.56) n=19	-0.22 (1.03) n=13	-0.23 (0.73) n=22	0.45 (1.19) n=19	-0.30 (0.99) n=28	0.52 (0.88) n=22	0.00 (1.11) n=22	**.000**	**.153**
Maths homework total scale	-0.37 (0.83) n=19	0.27 (0.74) n=20	-0.28 (1.09) n=13	-0.07 (0.65) n=22	0.30 (1.19) n=20	-0.23 (1.20) n=28	0.46 (0.89) n=22	0.02 (1.05) n=21	.063	.081
Maths environment	-0.08 (0.89) n=19	0.04 (0.80) n=20	0.29 (1.01) n=13	-0.08 (1.01) n=22	-0.34 (1.17) n=20	0.53 (0.68) n=28	0.10 (1.05) n=22	-0.85 (0.72) n=20	**.000**	**.166**
Maths usefulness	-0.07 (0.85) n=18	-0.05 (0.78) n=20	-0.44 (0.68) n=13	-0.18 (0.76) n=22	0.38 (1.24) n=20	0.03 (1.03) n=27	-0.59 (0.79) n=21	0.26 (1.09) n=18	.027	.098
Maths anxiety	-0.25 (0.87) n=19	-0.03 (0.84) n=19	-0.16 (1.07) n=14	0.20 (0.97) n=22	-0.14 (1.21) n=17	0.08 (1.08) n=26	0.16 (0.96) n=22	0.40 (1.18) n=17	.584	.037

Significant results in bold following a Bonferroni multiple testing correction of $p \leq .000$ ($p = .05$ divided by number of measures (k=114) across maths and geography, time 1, 2 & 3)

Table 5.2.7. Maths classroom variables at time 2 for school 2 (Russian sample): Means, standard deviation (SD), and N with ANOVA results by classroom, controlling for prior achievement

Construct	C9ce	C10ce	C11ce	*p*	η_p^2	Construct	C9ce	C10ce	C11ce	*p*	η_p^2
Maths Year 5 school achievement	-0.03 (0.69) n=14	0.18 (0.86) n=9	-1.05 (0.81) n=7	.009	.295						
Maths performance	0.36 (0.87) n=14	0.17 (0.84) n=9	-0.20 (0.89) n=6	.423	.064	Maths classroom chaos	-0.13 (1.06) n=14	0.90 (0.71) n=9	-0.59 (0.53) n=7	.005	.322
Number line	-1.06 (1.07) n=13	-0.37 (1.32) n=9	0.40 (0.60) n=7	.024	.250	Maths homework behaviour	0.09 (0.83) n=14	0.08 (0.69) n=9	0.48 (0.98) n=7	.556	.043
Maths self-perceived ability	-0.22 (0.81) n=14	-0.39 (0.69) n=9	-0.30 (1.09) n=7	.903	.008	Maths homework feedback	-0.11 (1.03) n=13	0.31 (0.54) n=9	-0.50 (1.16) n=7	.244	.103
Maths enjoyment	0.12 (1.06) n=14	-0.20 (0.52) n=8	-0.87 (1.39) n=7	.142	.140	Maths homework total scale	-0.06 (0.99) n=13	0.24 (0.52) n=9	-0.66 (1.34) n=7	.204	.115
Maths classroom environment	-0.94 (1.00) n=14	0.34 (0.47) n=9	-1.44 (1.08) n=7	.001	.398	Maths environment	0.17 (1.09) n=14	0.69 (1.00) n=8	-0.35 (0.92) n=7	.166	.129
Maths classroom student-teacher relations	-0.81 (1.07) n=14	0.35 (0.55) n=9	-1.26 (1.16) n=7	.006	.316	Maths usefulness	0.50 (1.20) n=14	-0.26 (0.55) n=9	0.93 (1.28) n=7	.089	.164
Maths classroom Peer competition	-0.91 (1.15) n=14	0.02 (0.75) n=9	-0.90 (1.19) n=6	.110	.156	Maths anxiety	-0.12 (0.76) n=14	-0.37 (0.80) n=9	-0.23 (1.16) n=7	.796	.017

Significant results in bold following a Bonferroni multiple testing correction of $p \leq .000$ ($p = .05$ divided by number of measures (k=114) across maths and geography, time 1, 2 & 3)

Table 5.2.8. Geography classroom variables at time 2 for school 1 (Russian sample): Means, standard deviation (SD), and N with ANOVA results by classroom, controlling for prior achievement

Construct	C1e	C2e	C3e	C4se	C5se	C6se	C7se	C8se	*p*	η_p^2
Geography Year 5 school achievement	0.54 (0.66) n=20	0.01 (1.02) n=20	0.32 (0.78) n=14	-0.37 (1.10) n=22	-0.65 (0.84) n=20	0.23 (0.75) n=28	0.22 (1.02) n=20	-0.23 (0.88) n=22	**.000**	**.151**
Geography performance	-0.30 (0.87) n=20	-0.34 (0.90) n=20	0.00 (0.86) n=14	-0.14 (1.19) n=21	-0.59 (0.92) n=19	0.46 (0.94) n=28	0.85 (0.77) n=21	-0.23 (1.08) n=21	**.000**	**.189**
Geography self-perceived ability	0.07 (0.88) n=19	-0.21 (0.79) n=20	0.53 (0.88) n=13	0.13 (0.98) n=21	-0.18 (1.16) n=18	-0.13 (1.08) n=27	0.42 (0.79) n=22	-0.78 (1.26) n=20	.004	.126
Geography enjoyment	0.34 (0.91) n=19	-0.14 (0.78) n=19	0.24 (0.92) n=14	0.21 (1.08) n=22	-0.12 (1.19) n=19	-0.15 (1.08) n=28	0.19 (0.86) n=22	-0.70 (1.08) n=19	.040	.090
Geography classroom environment	-0.07 (1.06) n=19	0.44 (0.69) n=20	0.53 (0.95) n=14	-0.31 (1.20) n=21	-0.43 (1.18) n=20	0.47 (0.77) n=28	0.03 (0.94) n=21	-0.26 (0.86) n=21	.003	.125
Geography classroom Student-teacher	-0.06 (1.04) n=19	0.44 (0.64) n=20	0.59 (0.88) n=14	-0.25 (1.18) n=21	-0.63 (1.17) n=20	0.42 (0.82) n=28	0.09 (0.92) n=21	-0.31 (0.82) n=21	**.000**	**.152**
Geography classroom peer competition	-0.11 (1.03) n=19	0.24 (0.80) n=20	0.18 (0.84) n=14	-0.32 (1.26) n=21	0.32 (1.03) n=19	0.37 (0.88) n=28	-0.13 (0.91) n=21	-0.05 (0.91) n=21	.197	.061

Significant results in bold following a Bonferroni multiple testing correction of $p \leq .000$ (p = .05 divided by number of measures (k=114) across maths and geography, time 1, 2 & 3)

Table 5.2.8. Continued. Geography classroom variables at time 2 for school 1 (Russian sample): Means, standard deviation (SD), and N with ANOVA results by classroom, controlling for prior achievement

Construct	C1e	C2e	C3e	C4se	C5se	C6se	C7se	C8se	*p*	η_p^2
Geography classroom chaos	-0.33 (0.99) n=19	0.15 (0.80) n=19	0.30 (1.22) n=14	-0.19 (1.10) n=21	-0.54 (0.88) n=20	0.00 (0.88) n=28	0.30 (0.78) n=20	-0.25 (0.92) n=21	.059	.084
Geography homework behaviour	-0.20 (1.04) n=19	0.28 (0.95) n=20	0.04 (1.03) n=14	-0.24 (0.88) n=21	0.11 (1.08) n=19	-0.06 (0.99) n=28	-0.36 (0.93) n=22	0.50 (1.06) n=21	.104	.072
Geography homework feedback	-0.76 (0.64) n=19	0.24 (0.69) n=20	0.30 (1.06) n=14	-0.44 (1.13) n=21	-0.16 (1.14) n=19	-0.03 (0.99) n=28	0.35 (1.11) n=22	0.22 (0.88) n=21	.003	.127
Geography homework total scale	-0.46 (0.73) n=19	0.03 (0.82) n=20	0.29 (1.18) n=14	-0.25 (0.96) n=21	-0.24 (1.18) n=19	-0.04 (1.05) n=28	0.43 (0.99) n=22	-0.09 (0.92) n=21	.112	.071
Geography environment	-0.19 (0.93) n=19	-0.22 (0.91) n=18	0.09 (1.03) n=13	-0.39 (0.67) n=20	-0.41 (0.97) n=18	0.96 (0.81) n=28	0.00 (0.98) n=21	-0.39 (1.06) n=20	**.000**	**.227**
Geography usefulness	0.17 (0.83) n=19	-0.03 (0.60) n=20	0.28 (0.80) n=14	-0.43 (1.28) n=22	-0.34 (1.22) n=20	0.10 (0.93) n=28	0.13 (1.10) n=22	-0.12 (0.86) n=20	.276	.053
Geography anxiety	-0.40 (0.96) n=19	-0.09 (0.93) n=20	-0.15 (1.24) n=13	0.37 (1.01) n=22	-0.08 (1.08) n=20	0.12 (0.93) n=27	-0.16 (0.78) n=22	0.48 (1.15) n=20	.111	.072

Significant results in bold following a Bonferroni multiple testing correction of $p \leq .000$ ($p = .05$ divided by number of measures (k=114) across maths and geography, time 1, 2 & 3)

Table 5.2.9. Geography classroom variables at time 2 for school 2 (Russian sample): Means, standard deviation (SD), and N with ANOVA results by classroom, controlling for prior achievement

Construct	C9ce	C10ce	C11ce	*p*	η_p^2	Construct	C9ce	C10ce	C11ce	*p*	η_p^2
Geography Year 5 school achievement	0.07 (0.84) n=14	0.91 (1.06) n=9	-0.01 (1.40) n=7	.134	.138	Geography classroom chaos	0.16 (1.06) n=14	1.28 (0.70) n=9	-0.09 (1.02) n=7	.013	.275
Geography performance	-0.20 (0.74) n=14	0.16 (0.76) n=9	0.40 (0.61) n=6	.215	.111	Geography homework behaviour	-0.34 (0.78) n=14	0.10 (0.74) n=8	0.47 (1.25) n=7	.162	.130
Geography self-perceived ability	0.29 (0.81) n=14	0.00 (0.63) n=9	0.32 (0.52) n=7	.560	.042	Geography homework feedback	0.44 (0.92) n=14	0.18 (0.67) n=8	-0.20 (0.85) n=7	.282	.093
Geography enjoyment	0.41 (0.92) n=14	-0.03 (0.80) n=9	-0.21 (0.42) n=7	.200	.112	Geography homework total scale	0.55 (0.89) n=14	0.27 (0.66) n=8	-0.37 (1.21) n=7	.118	.152
Geography classroom environment	-0.44 (0.91) n=14	0.01 (0.82) n=9	-0.31 (1.04) n=7	.522	.047	Geography environment	-0.17 (0.76) n=14	0.52 (0.74) n=8	0.10 (1.12) n=7	.207	.114
Geography classroom Student-teacher	-0.30 (0.96) n=14	0.13 (0.83) n=9	-0.34 (1.15) n=7	.533	.045	Geography usefulness	0.69 (0.86) n=14	-0.08 (0.96) n=9	-0.36 (1.07) n=7	.044	.207
Geography classroom peer competition	-0.58 (1.06) n=14	-0.26 (0.90) n=9	-0.11 (1.24) n=7	.593	.038	Geography anxiety	-0.19 (0.84) n=13	-0.17 (0.83) n=9	-0.08 (1.08) n=7	.965	.003

Significant results in bold following a Bonferroni multiple testing correction of $p \leq .000$ (p = .05 divided by number of measures (k=114) across maths and geography, time 1, 2 & 3)

Table 5.2.10. Maths teacher groups time 2 (Russian sample): Means, standard deviation (SD) and N for maths classroom variables with ANOVA results by teacher group, controlling for prior achievement

Construct	TM1	TM2	TM3	TM4	TM5	TM6	*p*	η_p^2
Maths Year 5 school achievement	-0.03 (0.69) n=14	-0.36 (1.03) n=16	0.36 (0.94) n=39	0.15 (0.74) n=14	0.40 (0.760 n=28	-0.23 (1.11) n=84	.005	.083
Maths performance	0.36 (0.87) n=14	0.02 (0.85) n=15	0.00 (1.08) n=39	0.13 (1.02) n=14	0.54 (0.73) n=28	-0.26 (1.01) n=86	.006	.082
Number line	-1.06 (1.07) n=13	-0.03 (1.11) n=16	0.20 (0.92) n=39	-0.19 (0.90) n=14	-0.11 (0.91) n=27	0.14 (0.96) n=85	.002	.098
Maths self-perceived ability	-0.22 (0.81) n=14	-0.35 (0.85) n=16	0.29 (0.91) n=38	-0.04 (1.01) n=13	0.15 (1.11) n=27	-0.07 (1.03) n=85	.216	.037
Maths enjoyment	0.12 (1.06) n=14	-0.51 (1.04) n=15	0.33 (0.88) n=37	-0.17 (1.10) n=13	-0.02 (1.11) n=28	-0.04 (0.95) n=86	.125	.045
Maths classroom environment	-0.94 (1.00) n=14	-0.44 (1.19) n=16	0.19 (0.76) n=39	0.38 (1.04) n=14	0.24 (0.86) n=28	0.01 (1.00) n=85	**.001**	**.105**
Maths classroom student-teacher relations	-0.81 (1.07) n=14	-0.35 (1.17) n=16	0.25 (0.73) n=39	0.32 (1.10) n=14	0.22 (0.920) n=28	-0.04 (0.99) n=85	.004	.085
Maths classroom peer competition	-0.91 (1.15) n=14	-0.35 (1.02) n=15	-0.17 (0.93) n=38	0.31 (0.75) n=14	0.12 (0.81) n=28	0.20 (1.00) n=85	**.001**	**.102**

Significant results in bold following a Bonferroni multiple testing correction of $p \leq .001$ ($p = .05$ divided by number of measures (k=57) across maths and geography, time 1, 2 & 3)

Table 5.2.10. Continued. Maths teacher groups time 2 (Russian sample): Means, standard deviation (SD) and N for maths classroom variables with ANOVA results by teacher group, controlling for prior achievement

Construct	TM1	TM2	TM3	TM4	TM5	TM6	*p*	η_p^2
Maths classroom chaos	-0.13 (1.06) n=14	0.26 (0.98) n=16	0.08 (0.78) n=39	0.53 (1.30) n=14	0.17 (0.79) n=28	-0.20 (1.05) n=86	.077	.050
Maths homework behaviour	0.09 (0.83) n=14	0.26 (0.83) n=16	0.06 (1.00) n=39	0.27 (1.01) n=13	-0.16 (1.06) n=28	-0.08 (1.03) n=86	.616	.018
Maths homework feedback	-0.11 (1.03) n=13	-0.05 (0.93) n=16	-0.03 (0.94) n=38	-0.22 (1.03) n=13	-0.30 (0.99) n=28	0.18 (1.02) n=85	.295	.032
Maths homework total scale	-0.06 (0.99) n=13	-0.15 (1.03) n=16	-0.04 (0.84) n=39	-0.28 (1.09) n=13	-0.23 (1.20) n=28	0.18 (0.97) n=85	.334	.030
Maths environment	0.17 (1.09) n=14	0.21 (1.07) n=15	-0.02 (0.84) n=39	0.29 (1.01) n=13	0.53 (0.68) n=28	-0.28 (1.05) n=84	.004	.087
Maths usefulness	0.50 (1.20) n=14	0.26 (1.09) n=16	-0.06 (0.81) n=38	-0.44 (0.68) n=13	0.03 (1.03) n=27	-0.05 (1.04) n=81	.190	.040
Maths anxiety	-0.12 (0.76) n=14	-0.31 (0.94) n=16	-0.14 (0.85) n=38	-0.16 (1.07) n=14	0.08 (1.08) n=26	0.16 (1.06) n=78	.426	.027

Significant results in bold following a Bonferroni multiple testing correction of $p \leq .001$ ($p = .05$ divided by number of measures (k=57) across maths and geography, time 1, 2 & 3)

Table 5.2.11. Geography teacher groups time 2 (Russian sample): Means, standard deviation (SD) and N for geography classroom variables with ANOVA results by teacher group, controlling for prior achievement

Construct	TG1	TG2	TG3	TG4	TG5	*p*	η_p^2
Geography Year 5 school achievement	0.30 (1.09) n=30	-0.22 (0.97) n=62	0.23 (0.75) n=28	0.28 (0.89) n=40	-0.10 (1.03) n=36	.027	.055
Geography performance	0.04 (0.74) n=29	0.03 (1.11) n=61	0.46 (0.94) n=28	-0.32 (0.87) n=40	-0.08 (1.06) n=35	.033	.054
Geography self-perceived ability	0.21 (0.69) n=30	-0.16 (1.17) n=60	-0.13 (1.08) n=27	-0.07 (0.84) n=39	0.28 (0.95) n=34	.175	.034
Geography enjoyment	0.14 (0.81) n=30	-0.19 (1.09) n=60	-0.15 (1.08) n=28	0.10 (0.87) n=38	0.22 (1.01) n=36	.238	.029
Geography classroom environment	-0.27 (0.91) n=30	-0.22 (1.00) n=62	0.47 (0.77) n=28	0.19 (0.91) n=39	0.03 (1.17) n=35	.011	.066
Geography classroom Student-teacher relations	-0.18 (0.96) n=30	-0.28 (1.00) n=62	0.42 (0.82) n=28	0.20 (0.88) n=39	0.08 (1.14) n=35	.013	.065
Geography classroom peer competition	-0.37 (1.04) n=30	0.04 (0.95) n=61	0.37 (0.88) n=28	0.07 (0.93) n=39	-0.12 (1.12) n=35	.064	.046

Significant results in bold following a Bonferroni multiple testing correction of $p \leq .001$ ($p = .05$ divided by number of measures (k=57) across maths and geography, time 1, 2 & 3)

Table 5.2.11. Continued. Geography teacher groups time 2 (Russian sample): Means, standard deviation (SD) and N for geography classroom variables with ANOVA results by teacher group, controlling for prior achievement

Construct	TG1	TG2	TG3	TG4	TG5	*p*	η_p^2
Geography classroom chaos	0.44 (1.09) n=30	-0.17 (0.92) n=61	0.00 (0.88) n=28	-0.09 (0.92) n=38	0.01 (1.16) n=35	.099	.041
Geography homework behaviour	-0.02 (0.93) n=29	0.08 (1.07) n=62	-0.06 (0.99) n=28	0.05 (1.01) n=39	-0.13 (0.94) n=35	.887	.006
Geography homework feedback	0.21 (0.85) n=29	0.15 (1.05) n=62	-0.03 (0.99) n=28	-0.25 (0.83) n=39	-0.14 (1.15) n=35	.209	.031
Geography homework total scale	0.25 (0.96) n=29	0.05 (1.05) n=62	-0.04 (1.05) n=28	-0.21 (0.81) n=39	-0.03 (1.07) n=35	.438	.020
Geography environment	0.09 (0.87) n=29	-0.26 (1.01) n=59	0.96 (0.81) n=28	-0.21 (0.91) n=37	-0.20 (0.85) n=33	**.000**	**.180**
Geography usefulness	0.21 (1.02) n=30	-0.10 (1.07) n=62	0.10 (0.93) n=28	0.06 (0.72) n=39	-0.15 (1.16) n=36	.522	.017
Geography anxiety	-0.16 (0.86) n=29	0.07 (1.03) n=62	0.12 (0.93) n=27	-0.24 (0.95) n=39	0.18 (1.12) n=35	.315	.025

Significant results in bold following a Bonferroni multiple testing correction of $p \leq .001$ ($p = .05$ divided by number of measures (k=57) across maths and geography, time 1, 2 & 3)

Table 5.2.12. Maths classroom variables at time 2 for school 1 (Russian sample): Classrooms ranked by means (highest = 1 to lowest = 8) for measures demonstrating a significant effect of maths classroom, controlling for prior achievement

Construct	C1e	C2e	C3e	C4se	C5se	C6se	C7se	C8se	*p*	η_p^2
Maths Year 5 school achievement	2nd	3rd	5th	7th	4th	1st	6th	8th	**.000**	**.151**
Maths performance*	4th	7th	3rd	2nd	6th	1st	5th	8th	.001	.142
Number line	5th	7th	2nd	1st	6th	3rd	4th	8th	**.000**	**.190**
Maths classroom chaos	5th	3rd	2nd	8th	7th	4th	1st	6th	**.000**	**.194**
Maths homework feedback	8th	2nd	5th	6th	3rd	7th	1st	4th	**.000**	**.153**
Maths environment	6th	4th	2nd	5th	7th	1st	3rd	8th	**.000**	**.166**

Significant results in bold following a Bonferroni multiple testing correction of $p \leq .000$ ($p = .05$ divided by number of measures (k=114) across maths and geography, time 1, 2 & 3). *Just below significance threshold, included for comparison.

Table 5.2.13. Maths classroom variables at time 2 for school 2 (Russian sample): Classrooms ranked by means (highest = 1 to lowest = 3) for measures demonstrating a significant effect of maths classroom, controlling for prior achievement

Construct	C9ce	C10ce	C11ce	*p*	η_p^2
Maths classroom Environment *	**2nd**	**1st**	**3rd**	.001	.398
Maths Year 5 school achievement*	**2nd**	**1st**	**3rd**	.009	.295

Significant results in bold following a Bonferroni multiple testing correction of $p \leq .000$(p = .05 divided by number of measures (k=114) across maths and geography, time 1, 2 & 3). *Just below significance threshold, included for comparison.

Table 5.2.14. Geography classroom variables at time 2 for school 1 (Russian sample): Classrooms ranked by means (highest = 1 to lowest = 8) for measures demonstrating a significant effect of geography classroom, controlling for prior achievement

Construct	C1e	C2e	C3e	C4se	C5se	C6se	C7se	C8se	*p*	η_p^2
Geography Year 5 school achievement	**1st**	**5th**	**2nd**	**7th**	**8th**	**3rd**	**4th**	**6th**	**.000**	**.151**
Geography performance	**6th**	**7th**	**3rd**	**4th**	**8th**	**2nd**	**1st**	**5th**	**.000**	**.189**
Geography classroom Student-teacher	**5th**	**2nd**	**1st**	**6th**	**8th**	**3rd**	**4th**	**7th**	**.000**	**.152**
Geography environment	**4th**	**5th**	**2nd**	**6th**	**8th**	**1st**	**3rd**	**7th**	**.000**	**.227**

Significant results in bold following a Bonferroni multiple testing correction of $p \leq .000$ ($p = .05$ divided by number of measures (k=114) across maths and geography, time 1, 2 & 3). *Not significant but ranked for comparison

Table 5.2.15. Maths Teacher groups at time 2 (Russian sample): ranked by means (highest = 1 to lowest = 6) for measures demonstrating a significant effect of maths teacher, controlling for prior achievement

Construct	TM1	TM2	TM3	TM4	TM5	TM6	*p*	η_p^2
Maths Year 5 school achievement*	4th	6th	2nd	3rd	1st	5th	.005	.083
Maths performance*	2nd	4th	5th	3rd	1st	6th	.006	.082
Number line*	1st	4th	6th	2nd	3rd	5th	**.002**	**.098**
Maths classroom environment	6th	5th	3rd	1st	2nd	4th	**.001**	**.105**
Maths classroom student-teacher relations*	6th	5th	2nd	1st	3rd	4th	.004	.085
Maths classroom peer competition	6th	5th	4th	1st	3rd	2nd	**.001**	**.102**

Significant results in bold following a Bonferroni multiple testing correction of $p \leq .001$ ($p = .05$ divided by number of measures (k=57) across maths and geography, time 1, 2 & 3). *Just below significance threshold, ranked for comparison.

Table 5.2.16. Geography Teacher groups at time 2 (Russian sample): ranked by means (highest = 1 to lowest = 5) for measures demonstrating a significant effect of geography teacher, controlling for prior achievement

Construct	TG1	TG2	TG3	TG4	TG5	*p*	η_p^2
Geography environment	**2nd**	**5th**	**1st**	**4th**	**3rd**	**.000**	**.180**
Geography Year 5 school achievement*	**1st**	**5th**	**3rd**	**2nd**	**4th**	.027	.055

Significant results in bold following a Bonferroni multiple testing correction of $p \leq .001$ ($p = .05$ divided by number of measures (k=57) across maths and geography, time 1, 2 & 3). *Not significant but ranked for comparison

Table 5.2.17. Maths classroom variables at time 3 for school 1 (Russian sample): Means, standard deviation (SD), and N with ANOVA results by classroom, controlling for prior achievement

Construct	C1e	C2e	C3e	C4se	C5se	C6se	C7se	C8se	*p*	η_p^2
Maths performance	0.52 (0.54) n=19	0.14 (0.90) n=18	0.06 (0.84) n=8	-0.11 (1.06) n=22	-0.59 (1.10) n=16	0.46 (0.83) n=28	-0.29 (1.03) n=19	-0.41 (1.14) n=20	.002	.147
Number line	-0.05 (0.81) n=21	0.02 (0.88) n=19	-0.45 (0.97) n=8	0.37 (0.81) n=17	-0.10 (0.72) n=28	-0.19 (0.99) n=18	0.68 (0.71) n=20	0.68 (0.71) n=20	.018	.109
Maths self-perceived ability	0.37 (0.98) n=21	-0.29 (0.71) n=19	0.23 (0.99) n=8	-0.24 (0.93) n=22	0.28 (1.16) n=17	0.05 (1.12) n=27	0.41 (1.09) n=19	-0.37 (0.97) n=20	.078	.083
Maths enjoyment	0.27 (0.97) n=20	0.03 (0.75) n=16	0.56 (1.01) n=8	-0.18 (0.83) n=22	-0.18 (0.81) n=16	0.08 (1.27) n=26	0.44 (0.94) n=19	-0.50 (1.04) n=20	.050	.095
Maths classroom environment	0.03 (0.98) n=20	0.50 (0.79) n=19	0.53 (0.88) n=8	0.02 (0.81) n=22	-0.38 (0.94) n=16	0.47 (0.85) n=28	-0.06 (0.77) n=19	-0.29 (1.05) n=20	.008	.123
Maths classroom Student-teacher relations	0.00 (0.94) n=20	0.65 (0.75) n=19	0.49 (0.98) n=8	-0.14 (0.79) n=22	-0.53 (0.88) n=16	0.50 (0.97) n=28	-0.03 (0.75) n=19	-0.21 (1.01) n=20	.001	.160
Maths classroom Peer competition	-0.01 (0.93) n=20	-0.14 (1.01) n=19	0.31 (0.62) n=8	0.50 (0.75) n=22	-0.01 (1.23) n=17	0.25 (0.77) n=28	-0.15 (1.04) n=19	-0.07 (0.95) n=20	.261	.059

Significant results in bold following a Bonferroni multiple testing correction of $p \leq .000$ (p = .05 divided by number of measures (k=114) across maths and geography, time 1, 2 & 3)

Table 5.2.17. Continued. Maths classroom variables at time 3 for school 1 (Russian sample): Means, standard deviation (SD), and N with ANOVA results by classroom, controlling for prior achievement

Construct	C1e	C2e	C3e	C4se	C5se	C6se	C7se	C8se	*p*	η_p^2
Maths classroom chaos	-0.17 (1.02) n=21	0.76 (0.87) n=19	-0.04 (1.45) n=8	-0.31 (0.88) n=22	-0.39 (0.86) n=17	0.23 (0.96) n=28	0.43 (0.97) n=19	-0.32 (0.95) n=21	.002	.144
Maths homework behaviour	-0.46 (0.86) n=20	0.37 (0.82) n=18	0.22 (0.41) n=8	0.33 (1.00) n=22	-0.39 (0.91) n=17	-0.37 (0.95) n=28	-0.34 (1.05) n=18	0.22 (1.27) n=21	.011	.118
Maths homework feedback	0.00 (0.97) n=20	0.46 (0.85) n=18	-0.42 (0.96) n=8	-0.39 (0.97) n=22	-0.41 (0.83) n=17	0.03 (1.20) n=28	0.09 (0.96) n=17	0.18 (1.13) n=21	.120	.076
Maths homework total scale	0.24 (0.71) n=20	0.14 (0.88) n=18	-0.37 (0.90) n=8	-0.40 (0.92) n=21	-0.12 (0.93) n=17	0.17 (1.26) n=28	0.22 (1.11) n=18	0.06 (1.13) n=20	.378	.051
Maths environment	-0.10 (1.06) n=20	0.54 (0.67) n=19	-0.07 (0.73) n=8	-0.12 (1.04) n=22	-0.06 (1.05) n=17	0.43 (0.97) n=27	-0.12 (1.01) n=19	-0.55 (1.06) n=19	.017	.111
Maths usefulness	-0.29 (0.80) n=20	-0.14 (0.77) n=19	0.19 (1.08) n=6	-0.10 (0.90) n=21	0.16 (1.12) n=17	-0.11 (0.98) n=27	-0.74 (0.82) n=19	0.39 (1.09) n=21	.021	.107
Maths anxiety	-0.28 (0.71) n=20	0.14 (0.72) n=19	-0.58 (0.65) n=8	0.46 (1.02) n=22	-0.32 (0.99) n=17	-0.06 (1.03) n=27	0.01 (1.12) n=19	0.52 (1.18) n=20	.020	.107

Significant results in bold following a Bonferroni multiple testing correction of $p \leq .000$ (p = .05 divided by number of measures (k=114) across maths and geography, time 1, 2 & 3)

Table 5.2.18. Maths classroom variables at time 3 for school 2 (Russian sample): Means, standard deviation (SD), and N with ANOVA results by classroom controlling for prior achievement

Construct	C9ce	C10ce	C11ce	*p*	η_p^2	Construct	C9ce	C10ce	C11ce	*p*	η_p^2
Maths performance	0.23 (0.93) n=15	0.42 (0.78) n=10	-0.38 (1.09) n=7	.214	.101	Maths classroom chaos	-0.12 (0.85) n=15	0.71 (0.40) n=10	-0.59 (0.49) n=7	.001	.367
Number line	-0.56 (1.07) n=14	-0.15 (0.79) n=10	1.07 (1.34) n=7	.009	.287	Maths homework behaviour	-0.03 (0.84) n=15	0.19 (0.67) n=10	1.04 (0.68) n=7	.015	.251
Maths self-perceived ability	-0.41 (0.83) n=15	0.22 (0.78) n=9	0.37 (1.15) n=6	.119	.146	Maths homework feedback	-0.07 (0.66) n=15	0.23 (0.76) n=10	0.42 (1.30) n=7	.437	.055
Maths enjoyment	-0.14 (1.14) n=15	-0.20 (0.81) n=8	0.22 (0.99) n=6	.724	.024	Maths homework total scale	0.00 (0.78) n=15	0.13 (0.68) n=10	-0.39 (1.20) n=7	.464	.052
Maths classroom environment	-0.43 (0.96) n=15	0.52 (0.79) n=10	-1.94 (1.03) n=6	**.000**	**.489**	Maths environment	0.24 (0.93) n=14	0.17 (0.93) n=10	-0.48 (0.56) n=6	.234	.102
Maths classroom student-teacher relations	-0.44 (0.99) n=15	0.36 (0.73) n=10	-1.55 (1.27) n=6	.003	.338	Maths usefulness	0.12 (0.99) n=15	0.44 (0.74) n=10	0.99 (1.25) n=7	.170	.115
Maths classroom Peer competition	-0.25 (1.03) n=15	0.32 (0.69) n=10	-1.45 (1.14) n=6	.005	.317	Maths anxiety	-0.05 (1.02) n=15	-0.71 (0.71) n=10	-0.02 (0.89) n=6	.173	.118

Significant results in bold following a Bonferroni multiple testing correction of $p \leq .000$ (p = .05 divided by number of measures (k=114) across maths and geography, time 1, 2 & 3)

Table 5.2.19. Geography classroom variables at time 3 for school 1 (Russian sample): Means, standard deviation (SD), and N with ANOVA results by classroom, controlling for prior achievement

Construct	C1e	C2e	C3e	C4se	C5se	C6se	C7se	C8se	*p*	η_p^2
Geography performance	-0.04 (0.88) n=22	-0.15 (0.85) n=19	0.12 (0.98) n=8	-0.17 (1.13) n=22	-0.50 (1.06) n=16	0.47 (0.83) n=28	0.60 (0.82) n=19	-0.08 (1.07) n=20	.008	.120
Geography self-perceived ability	0.12 (0.99) n=22	-0.12 (0.78) n=19	0.27 (1.25) n=6	-0.33 (1.00) n=22	0.08 (1.08) n=17	-0.10 (1.00) n=27	0.61 (0.82) n=19	-0.29 (1.26) n=21	.097	.079
Geography enjoyment	0.40 (0.93) n=18	-0.02 (0.65) n=19	0.13 (0.78) n=6	-0.06 (0.85) n=22	-0.27 (1.33) n=17	0.04 (0.90) n=27	0.29 (0.88) n=19	-0.61 (1.26) n=21	.059	.090
Geography classroom environment	-0.32 (1.10) n=22	0.25 (0.88) n=19	0.09 (1.05) n=6	-0.04 (1.19) n=22	-0.36 (1.02) n=17	0.79 (0.54) n=27	0.27 (0.71) n=18	-0.35 (0.91) n=21	**.000**	**.172**
Geography classroom Student-teacher	-0.23 (1.01) n=21	0.38 (0.76) n=19	0.41 (0.97) n=6	0.09 (1.08) n=22	-0.38 (1.08) n=17	0.67 (0.71) n=27	0.12 (1.12) n=19	-0.42 (0.92) n=21	.002	.146
Geography classroom peer competition	-0.26 (1.20) n=21	0.14 (1.08) n=18	-0.25 (0.90) n=6	0.07 (1.17) n=22	0.05 (0.80) n=17	0.35 (0.84) n=27	-0.12 (0.86) n=19	0.04 (1.02) n=21	.577	.038

Significant results in bold following a Bonferroni multiple testing correction of $p \leq .000$ ($p = .05$ divided by number of measures (k=114) across maths and geography, time 1, 2 & 3)

Table 5.2.19. Continued. Geography classroom variables at time 3 for school 1 (Russian sample): Means, standard deviation (SD), and N with ANOVA results by classroom, controlling for prior achievement

Construct	C1e	C2e	C3e	C4se	C5se	C6se	C7se	C8se	*p*	η_p^2
Geography classroom chaos	-0.08 (1.00) n=21	0.40 (0.98) n=19	0.04 (1.55) n=6	-0.18 (1.13) n=21	-0.82 (1.11) n=17	0.24 (0.79) n=26	0.32 (0.86) n=19	-0.22 (0.80) n=21	.007	.125
Geography homework behaviour	-0.24 (0.83) n=18	0.30 (0.97) n=18	-0.24 (0.79) n=6	0.02 (1.02) n=21	-0.13 (0.85) n=17	-0.03 (0.97) n=27	-0.32 (0.86) n=19	0.38 (1.28) n=21	.278	.059
Geography homework feedback	-0.34 (0.87) n=19	0.24 (0.77) n=18	-0.09 (1.27) n=6	-0.10 (1.05) n=21	-0.20 (1.03) n=17	0.30 (0.86) n=27	0.18 (1.16) n=19	-0.29 (1.35) n=21	.317	.056
Geography homework total scale	-0.22 (0.93) n=19	-0.03 (0.85) n=18	0.03 (1.19) n=6	-0.14 (0.82) n=21	-0.14 (0.81) n=17	0.25 (1.03) n=27	0.30 (1.15) n=19	-0.29 (1.31) n=20	.483	.045
Geography environment	-0.22 (0.89) n=18	0.18 (0.79) n=19	0.08 (1.18) n=6	-0.39 (0.81) n=20	-0.31 (1.10) n=17	0.95 (0.82) n=27	-0.37 (0.83) n=19	-0.56 (0.94) n=21	**.000**	**.262**
Geography usefulness	-0.11 (0.89) n=20	-0.08 (0.97) n=19	0.15 (2.04) n=4	-0.01 (0.85) n=20	-0.26 (1.04) n=17	0.17 (1.15) n=27	0.15 (1.25) n=19	-0.24 (1.05) n=20	.847	.024
Geography anxiety	-0.07 (0.88) n=21	0.26 (0.90) n=19	-0.34 (0.95) n=6	0.91 (1.21) n=22	-0.16 (1.02) n=16	-0.05 (1.02) n=27	-0.33 (0.72) n=19	-0.04 (1.06) n=21	.004	.135

Significant results in bold following a Bonferroni multiple testing correction of $p \leq .000$ ($p = .05$ divided by number of measures (k=114) across maths and geography, time 1, 2 & 3)

Table 5.2.20. Geography classroom variables at time 3 for school 2 (Russian sample): Means, standard deviation (SD), and N with ANOVA results by classroom, controlling for prior achievement

Construct	C9ce	C10ce	C11ce	*p*	η_p^2	Construct	C9ce	C10ce	C11ce	*p*	η_p^2
Geography performance	0.00 (0.93) n=15	0.28 (0.91) n=10	-0.96 (0.68) n=6	.034	.215						
Geography self-perceived ability	-0.03 (0.75) n=15	0.28 (0.77) n=10	0.13 (1.18) n=6	.665	.029	Geography homework behaviour	-0.27 (1.11) n=15	-0.18 (0.81) n=8	0.86 (0.84) n=6	.069	.186
Geography enjoyment	0.30 (0.91) n=15	0.09 (0.85) n=9	0.39 (1.20) n=5	.814	.016	Geography homework feedback	0.13 (0.90) n=15	0.09 (0.79) n=8	0.21 (0.75) n=5	.972	.002
Geography classroom environment	-0.56 (0.96) n=15	0.24 (0.95) n=10	-0.71 (0.89) n=7	.073	.165	Geography homework total scale	0.21 (1.10) n=15	0.29 (0.74) n=8	-0.07 (0.86) n=5	.799	.018
Geography classroom Student-teacher	-0.64 (0.89) n=15	0.11 (0.90) n=10	-0.36 (0.85) n=7	.134	.130	Geography environment	-0.07 (0.96) n=15	0.68 (0.64) n=10	0.07 (0.76) n=6	.099	.152
Geography classroom peer competition	-0.34 (0.97) n=15	0.13 (0.96) n=10	-0.70 (1.04) n=7	.233	.096	Geography usefulness	0.25 (0.73) n=15	-0.07 (0.73) n=10	0.39 (0.71) n=6	.407	.062
Geography classroom chaos	0.21 (0.73) n=15	0.89 (0.44) n=10	-0.51 (0.84) n=7	.001	.377	Geography anxiety	-0.29 (0.89) n=15	-0.32 (0.53) n=10	-0.10 (0.80) n=6	.841	.012

Significant results in bold following a Bonferroni multiple testing correction of *p* ≤ .000 (*p* = .05 divided by number of measures (k=114) across maths and geography, time 1, 2 & 3)

Table 5.2.21. Maths teacher groups time 3 (Russian sample): Means, standard deviation (SD) and N for maths classroom variables with ANOVA results by teacher group, controlling for prior achievement

Construct	**TM1**	**TM2**	**TM3**	**TM4**	**TM5**	**TM6**	***p***	**η_p^2**
Maths performance	0.23 (0.93) n=15	0.09 (0.97) n=17	0.33 (0.75) n=37	0.06 (0.84) n=8	0.46 (0.83) n=28	-0.33 (1.08) n=77	**.001**	**.108**
Number line	-0.56 (1.07) n=14	0.35 (1.19) n=17	-0.02 (0.83) n=40	-0.45 (0.97) n=8	-0.10 (0.72) n=28	0.17 (1.06) n=76	.056	.059
Maths self-perceived ability	-0.41 (0.83) n=15	0.28 (0.91) n=15	0.06 (0.91) n=40	0.23 (0.99) n=8	0.05 (1.12) n=27	0.00 (1.06) n=78	.531	.023
Maths enjoyment	-0.14 (1.14) n=15	-0.02 (0.88) n=14	0.16 (0.88) n=36	0.56 (1.01) n=8	0.08 (1.27) n=26	-0.11 (0.96) n=77	.445	.027
Maths classroom environment	-0.43 (0.96) n=15	-0.40 (1.50) n=16	0.26 (0.91) n=39	0.53 (0.88) n=8	0.47 (0.85) n=28	-0.16 (0.89) n=77	.002	.101
Maths classroom student-teacher relations	-0.44 (0.99) n=15	-0.36 (1.33) n=16	0.32 (0.90) n=39	0.49 (0.98) n=8	0.50 (0.97) n=28	-0.21 (0.86) n=77	**.000**	**.118**
Maths classroom peer competition	-0.25 (1.03) n=15	-0.35 (1.22) n=16	-0.08 (0.96) n=39	0.31 (0.62) n=8	0.25 (0.77) n=28	0.09 (1.01) n=78	.276	.035

Significant results in bold following a Bonferroni multiple testing correction of $p \leq .001$ ($p = .05$ divided by number of measures (k=57) across maths and geography, time 1, 2 & 3)

Table 5.2.21. Continued. Maths teacher groups time 3 (Russian sample): Means, standard deviation (SD) and N for maths classroom variables with ANOVA results by teacher group, controlling for prior achievement

Construct	TM1	TM2	TM3	TM4	TM5	TM6	*p*	η_p^2
Maths classroom chaos	-0.12 (0.85) n=15	0.17 (0.78) n=17	0.27 (1.05) n=40	-0.04 (1.45) n=8	0.23 (0.96) n=28	-0.15 (0.96) n=79	.233	.037
Maths homework behaviour	-0.03 (0.84) n=15	0.54 (0.78) n=17	-0.07 (0.93) n=38	0.22 (0.41) n=8	-0.37 (0.95) n=28	-0.01 (1.10) n=78	.085	.052
Maths homework feedback	-0.07 (0.66) n=15	0.30 (0.98) n=17	0.21 (0.93) n=38	-0.42 (0.96) n=8	0.03 (1.20) n=28	-0.13 (1.00) n=77	.288	.034
Maths homework total scale	0.00 (0.78) n=15	-0.09 (0.93) n=17	0.19 (0.78) n=38	-0.37 (0.90) n=8	0.17 (1.26) n=28	-0.07 (1.04) n=76	.584	.021
Maths environment	0.24 (0.93) n=14	-0.07 (0.85) n=16	0.21 (0.94) n=39	-0.07 (0.73) n=8	0.43 (0.97) n=27	-0.21 (1.04) n=77	.045	.062
Maths usefulness	0.12 (0.99) n=15	0.66 (0.99) n=17	-0.22 (0.78) n=39	0.19 (1.08) n=6	-0.11 (0.98) n=27	-0.07 (1.06) n=78	.059	.058
Maths anxiety	-0.05 (1.02) n=15	-0.45 (0.83) n=16	-0.07 (0.73) n=39	-0.58 (0.65) n=8	-0.06 (1.03) n=27	0.19 (1.11) n=78	.094	.051

Significant results in bold following a Bonferroni multiple testing correction of $p \leq .001$ ($p = .05$ divided by number of measures (k=57) across maths and geography, time 1, 2 & 3)

Table 5.2.22. Geography teacher groups time 3 (Russian sample): Means, standard deviation (SD) and N for geography classroom variables with ANOVA results by teacher group, controlling for prior achievement

Construct	TG1	TG2	TG3	TG4	TG5	*p*	η_p^2
Geography performance	-0.09 (0.97) n=31	0.03 (1.07) n=55	0.47 (0.83) n=28	-0.09 (0.86) n=41	-0.10 (1.08) n=30	.121	.040
Geography self-perceived ability	0.10 (0.83) n=31	0.12 (1.12) n=57	-0.10 (1.00) n=27	0.01 (0.89) n=41	-0.20 (1.07) n=28	.625	.014
Geography enjoyment	0.25 (0.92) n=29	-0.21 (1.21) n=57	0.04 (0.90) n=27	0.18 (0.81) n=37	-0.02 (0.82) n=28	.235	.031
Geography classroom environment	-0.34 (1.00) n=32	-0.15 (0.92) n=56	0.79 (0.54) n=27	-0.06 (1.03) n=41	-0.01 (1.14) n=28	**.000**	**.119**
Geography classroom Student-teacher relations	-0.35 (0.92) n=32	-0.23 (1.05) n=57	0.67 (0.71) n=27	0.06 (0.94) n=40	0.16 (1.05) n=28	**.000**	**.107**
Geography classroom peer competition	-0.27 (1.00) n=32	-0.01 (0.89) n=57	0.35 (0.84) n=27	-0.07 (1.15) n=39	0.00 (1.11) n=28	.206	.033

Significant results in bold following a Bonferroni multiple testing correction of $p \leq .001$ ($p = .05$ divided by number of measures (k=57) across maths and geography, time 1, 2 & 3)

Table 5.2.22. Continued. Geography teacher groups time 3 (Russian sample): Means, standard deviation (SD) and N for geography classroom variables with ANOVA results by teacher group, controlling for prior achievement

Construct	TG1	TG2	TG3	TG4	TG5	*p*	η_p^2
Geography classroom chaos	0.27 (0.83) n=32	-0.22 (1.01) n=57	0.24 (0.79) n=26	0.14 (1.01) n=40	-0.13 (1.21) n=27	.101	.043
Geography homework behaviour	-0.01 (1.05) n=29	-0.01 (1.06) n=57	-0.03 (0.97) n=27	0.03 (0.93) n=36	-0.04 (0.97) n=27	.999	.000
Geography homework feedback	0.13 (0.81) n=28	-0.11 (1.19) n=57	0.30 (0.86) n=27	-0.06 (0.86) n=37	-0.10 (1.08) n=27	.425	.022
Geography homework total scale	0.18 (0.95) n=28	-0.05 (1.13) n=56	0.25 (1.03) n=27	-0.13 (0.88) n=37	-0.10 (0.89) n=27	.471	.021
Geography environment	0.20 (0.88) n=31	-0.42 (0.95) n=57	0.95 (0.82) n=27	-0.01 (0.85) n=37	-0.28 (0.91) n=26	**.000**	**.216**
Geography usefulness	0.18 (0.72) n=31	-0.11 (1.11) n=56	0.17 (1.15) n=27	-0.10 (0.91) n=39	0.01 (1.07) n=24	.597	.016
Geography anxiety	-0.26 (0.76) n=31	-0.17 (0.94) n=56	-0.05 (1.02) n=27	0.08 (0.90) n=40	0.64 (1.25) n=28	.003	.086

Significant results in bold following a Bonferroni multiple testing correction of $p \leq .001$ ($p = .05$ divided by number of measures (k=57) across maths and geography, time 1, 2 & 3)

Table 5.2.23. Maths classroom variables at time 3 for school 1 (Russian sample): Classrooms ranked by means (highest = 1 to lowest = 8) for measures demonstrating a significant effect of maths classroom, controlling for prior achievement

Construct	C1e	C2e	C3e	C4se	C5se	C6se	C7se	C8se	*p*	η_p^2
Maths performance	**1st**	**3rd**	**4th**	**5th**	**8th**	**2nd**	**6th**	**7th**	.002	.147
Maths classroom environment	**4th**	**2nd**	**1st**	**5th**	**8th**	**3rd**	**6th**	**7th**	.008	.123
Maths classroom Student-teacher relations	**4th**	**1st**	**3rd**	**6th**	**8th**	**2nd**	**5th**	**7th**	.001	.160
Maths classroom chaos	**5th**	**1st**	**4th**	**6th**	**8th**	**3rd**	**2nd**	**7th**	.002	.144

Significant results in bold following a Bonferroni multiple testing correction of $p \leq .000$ ($p = .05$ divided by number of measures (k=114) across maths and geography, time 1, 2 & 3)

Table 5.2.24. Maths classroom variables at time 3 for school 2 (Russian sample): Classrooms ranked by means (highest = 1 to lowest = 3) for measures demonstrating a significant effect of maths classroom, controlling for prior achievement

Construct	C9ce	C10ce	C11ce	p	η_p^2
Maths classroom environment	**2nd**	**1st**	**3rd**	**.000**	**.489**
Maths classroom student-teacher relations *	**2nd**	**1st**	**3rd**	.003	.338
Maths classroom Peer competition*	**2nd**	**1st**	**3rd**	.005	.317
Maths classroom Chaos*	**2nd**	**1st**	**3rd**	.001	.367

Significant results in bold following a Bonferroni multiple testing correction of $p \leq .000$($p = .05$ divided by number of measures (k=114) across maths and geography, time 1, 2 & 3). *Just below significance but ranked for comparison.

Table 5.2.25. Geography classroom variables at time 3 for school 1 (Russian sample): Classrooms ranked by means (highest = 1 to lowest = 8) for measures demonstrating a significant effect of geography classroom at time 2, controlling for prior achievement

Construct	C1e	C2e	C3e	C4se	C5se	C6se	C7se	C8se	*p*	η_p^2
Geography performance*	4th	6th	3rd	7th	8th	2nd	1st	5th	.008	.120
Geography classroom environment	6th	3rd	4th	5th	8th	1st	2nd	7th	**.000**	**.172**
Geography classroom Student-teacher*	6th	3rd	2nd	5th	7th	1st	4th	8th	.002	.146
Geography classroom chaos	5th	1st	4th	6th	8th	3rd	2nd	7th	.007	.125
Geography environment	4th	2nd	3rd	7th	5th	1st	6th	8th	**.000**	**.262**
Geography anxiety*	5th	2nd	8th	1st	6th	4th	7th	3rd	.004	.135

Significant results in bold following a Bonferroni multiple testing correction of $p \leq .000$ ($p = .05$ divided by number of measures (k=114) across maths and geography, time 1, 2 & 3). *Just below significance but ranked for comparison. Anxiety: high score = high anxiety.

Table 5.2.26. Geography classroom variables at time 3 for school 2 (Russian sample): Classrooms ranked by means (highest = 1 to lowest = 3) for measures demonstrating a significant effect of geography classroom, controlling for prior achievement

Construct	C9ce	C10ce	C11ce	p	η_p^2
Geography performance	2nd	1st	3rd	.034	.215
Geography classroom Chaos	2nd	1st	3rd	.001	.377

Significant results in bold following a Bonferroni multiple testing correction of $p \leq .000$ (p = .05 divided by number of measures (k=114) across maths and geography, time 1, 2 & 3)

Table 5.2.27. Maths Teacher groups at time 3 (Russian sample): ranked by means (highest = 1 to lowest = 6) for measures demonstrating a significant effect of maths teacher, controlling for prior achievement

Construct	TM1	TM2	TM3	TM4	TM5	TM6	p	η_p^2
Maths performance	3rd	4th	2nd	5th	1st	6th	**.001**	**.108**
Maths classroom environment*	6th	5th	3rd	1st	2nd	4th	.002	.101
Maths classroom student-teacher relations	6th	5th	3rd	2nd	1st	4th	**.000**	**.118**

Significant results in bold following a Bonferroni multiple testing correction of $p \leq .001$ (p = .05 divided by number of measures (k=57) across maths and geography, time 1, 2 & 3). *Just below significance but ranked for comparison.

Table 5.2.28. Geography Teacher groups at time 3 (Russian sample): ranked by means (highest = 1 to lowest = 5) for measures demonstrating a significant effect of geography teacher controlling for prior achievement

Construct	TG1	TG2	TG3	TG4	TG5	*p*	η_p^2
Geography classroom environment	**5th**	**4th**	**1st**	**3rd**	**2nd**	**.000**	**.119**
Geography classroom Student-teacher relations	**5th**	**4th**	**1st**	**3rd**	**2nd**	**.000**	**.107**
Geography environment	**2nd**	**5th**	**1st**	**3rd**	**4th**	**.000**	**.216**
Geography anxiety*	**5th**	**4th**	**3rd**	**2nd**	**1st**	.003	.086

Significant results in bold following a Bonferroni multiple testing correction of $p \leq .001$ ($p = .05$ divided by number of measures (k=57) across maths and geography, time 1, 2 & 3). *Just below significance but ranked for comparison. Anxiety: high score = high anxiety

Table 5.2.29. Levene's test of equality of variances for school 1 maths classroom measures controlling for prior achievement at time 2 and time 3

Wave	Construct	F	df1	df2	Sig.
Time 1	Maths performance	.803	7	164	.586
Time 2	Maths Year 5 achievement	1.424	7	157	.199
	Maths performance	1.023	7	159	.417
	Number line	1.264	7	157	.272
	Maths self-perceived ability	1.288	7	155	.259
	Maths enjoyment	.797	7	156	.591
	Maths classroom environment	1.679	7	158	.118
	Maths student-teacher relations	1.383	7	158	.216
	Maths peer competition	.688	7	157	.682
	Maths classroom chaos	2.772	7	159	**.010**
	Maths homework behaviour	.677	7	158	.691
	Maths homework feedback	2.154	7	156	**.041**
	Maths homework total scale	1.422	7	157	.200
	Maths environment	2.116	7	156	**.045**
	Maths usefulness	1.120	7	151	.354
	Maths anxiety	1.318	7	148	.246
Time 3	Maths performance	1.580	7	142	.146
	Number line	3.607	7	144	**.001**
	Maths self-perceived ability	.532	7	145	.809
	Maths enjoyment	1.435	7	139	.196
	Maths classroom environment	.820	7	144	.572
	Maths student-teacher relations	.922	7	144	.492
	Maths peer competition	1.969	7	145	.063
	Maths classroom chaos	.694	7	147	.677
	Maths homework behaviour	2.277	7	144	**.031**
	Maths homework feedback	.825	7	143	.568
	Maths homework total scale	1.205	7	142	.304
	Maths environment	.704	7	143	.669
	Maths usefulness	.495	7	142	.837
	Maths anxiety	2.400	7	144	**.024**

Bold = significant at $p \leq .05$

Table 5.2.30. Levene's test of equality of variances for school 2 maths classroom measures controlling for prior achievement at time 2 and time 3

Wave	Construct	F	df1	df2	Sig.
Time 1	Maths performance	.906	2	32	.414
Time 2	Maths Year 5 achievement	.814	2	27	.454
	Maths performance	.005	2	26	.995
	Number line	1.912	2	26	.168
	Maths self-perceived ability	1.796	2	27	.185
	Maths enjoyment	1.832	2	26	.180
	Maths classroom environment	3.270	2	27	.054
	Maths student-teacher relations	3.270	2	27	.054
	Maths peer competition	.704	2	26	.504
	Maths classroom chaos	2.474	2	27	.103
	Maths homework behaviour	.739	2	27	.487
	Maths homework feedback	2.772	2	26	.081
	Maths homework total scale	3.177	2	26	.058
	Maths environment	.245	2	26	.785
	Maths usefulness	1.530	2	27	.235
	Maths anxiety	1.406	2	27	.262
Time 3	Maths performance	.342	2	29	.714
	Number line	1.042	2	28	.366
	Maths self-perceived ability	.692	2	27	.509
	Maths enjoyment	.392	2	26	.680
	Maths classroom environment	.785	2	28	.466
	Maths student-teacher relations	1.052	2	28	.363
	Maths peer competition	1.047	2	28	.364
	Maths classroom chaos	7.001	2	29	**.003**
	Maths homework behaviour	.333	2	29	.720
	Maths homework feedback	1.546	2	29	.230
	Maths homework total scale	1.038	2	29	.367
	Maths environment	.475	2	27	.627
	Maths usefulness	.852	2	29	.437
	Maths anxiety	2.037	2	28	.149

Bold = significant at $p \leq .05$

Table 5.2.31. Levene's test of equality of variances for school 1 geography classroom measures controlling for prior achievement at time 2 and time 3

Wave	Construct	F	df1	df2	Sig.
Time 2	Geography Year 5 achievement	1.379	7	158	.218
	Geography performance	.522	7	156	.817
	Geography self-perceived ability	1.117	7	152	.355
	Geography enjoyment	.611	7	154	.746
	Geography classroom environment	1.488	7	156	.175
	Geography S-T relations	1.849	7	156	.082
	Geography peer competition	1.888	7	155	.075
	Geography classroom chaos	1.311	7	154	.249
	Geography homework behaviour	.620	7	156	.739
	Geography homework feedback	2.433	7	156	**.022**
	Geography homework total scale	1.176	7	156	.320
	Geography environment	.824	7	149	.569
	Geography usefulness	1.898	7	157	.073
	Geography anxiety	.622	7	155	.737
Time 3	Geography performance	.552	7	146	.794
	Geography self-perceived ability	.547	7	145	.798
	Geography enjoyment	1.950	7	141	.066
	Geography classroom environment	3.094	7	144	**.005**
	Geography S-T relations	1.239	7	144	.285
	Geography peer competition	2.052	7	143	.053
	Geography classroom chaos	2.210	7	142	**.037**
	Geography homework behaviour	1.087	7	139	.375
	Geography homework feedback	1.613	7	140	.136
	Geography homework total scale	1.717	7	139	.110
	Geography environment	.897	7	139	.511
	Geography usefulness	.923	7	138	.491
	Geography anxiety	1.028	7	143	.414

Bold = significant at $p \leq .05$. **S-T** =student-teacher

Table 5.2.32. Levene's test of equality of variances for school 2 geography classroom measures controlling for prior achievement at time 2 and time 3

Wave	Construct	F	df1	df2	Sig.
Time 2	Geography Year 5 achievement	1.244	2	27	.304
	Geography performance	.294	2	26	.748
	Geography self-perceived ability	1.796	2	27	.185
	Geography enjoyment	1.795	2	27	.185
	Geography classroom environment	.525	2	27	.597
	Geography S-T relations	.928	2	27	.408
	Geography peer competition	.757	2	27	.479
	Geography classroom chaos	2.691	2	27	.086
	Geography homework behaviour	3.201	2	26	.057
	Geography homework feedback	.656	2	26	.528
	Geography homework total scale	1.961	2	26	.161
	Geography environment	1.311	2	26	.287
	Geography usefulness	.014	2	27	.986
	Geography anxiety	.490	2	26	.618
Time 3	Geography performance	.605	2	28	.553
	Geography self-perceived ability	1.560	2	28	.228
	Geography enjoyment	.108	2	26	.898
	Geography classroom environment	.194	2	29	.825
	Geography S-T relations	.024	2	29	.977
	Geography peer competition	.195	2	29	.824
	Geography classroom chaos	2.316	2	29	.117
	Geography homework behaviour	1.691	2	26	.204
	Geography homework feedback	.270	2	25	.766
	Geography homework total scale	1.314	2	25	.287
	Geography environment	.811	2	28	.454
	Geography usefulness	.046	2	28	.955
	Geography anxiety	1.141	2	28	.334

Bold = significant at $p \leq .05$. **S-T** =student-teacher

Table 5.2.33. Levene's test of equality of variances for maths classroom measures by maths teacher controlling for prior achievement at time 2 and time 3

Wave	Construct	F	df1	df2	Sig.
Time 1	Maths performance	1.521	5	201	.185
Time 2	Maths Year 5 achievement	1.810	5	189	.113
	Maths performance	1.464	5	190	.203
	Number line	.475	5	188	.794
	Maths self-perceived ability	.529	5	187	.754
	Maths enjoyment	.165	5	187	.975
	Maths classroom environment	1.686	5	190	.140
	Maths student-teacher relations	1.722	5	190	.131
	Maths peer competition	.507	5	188	.771
	Maths classroom chaos	3.051	5	191	**.011**
	Maths homework behaviour	1.085	5	190	.370
	Maths homework feedback	.203	5	187	.961
	Maths homework total scale	.472	5	188	.797
	Maths environment	1.878	5	187	.100
	Maths usefulness	.974	5	183	.435
	Maths anxiety	1.650	5	180	.149
Time 3	Maths performance	1.683	5	176	.141
	Number line	2.498	5	177	**.033**
	Maths self-perceived ability	.477	5	177	.793
	Maths enjoyment	1.433	5	170	.215
	Maths classroom environment	2.310	5	177	**.046**
	Maths student-teacher relations	1.558	5	177	.174
	Maths peer competition	.996	5	178	.421
	Maths classroom chaos	1.309	5	181	.262
	Maths homework behaviour	2.660	5	178	**.024**
	Maths homework feedback	1.296	5	177	.268
	Maths homework total scale	1.400	5	176	.226
	Maths environment	.484	5	175	.788
	Maths usefulness	.661	5	176	.653
	Maths anxiety	3.053	5	177	**.011**

Bold = significant at $p \leq .05$

Table 5.2.34. Levene's test of equality of variances for geography classroom measures by geography teacher controlling for prior achievement at time 2 and time 3

Wave	Construct	F	df1	df2	Sig.
Time 2	Geography Year 5 achievement	.895	4	191	.468
	Geography performance	1.309	4	188	.268
	Geography self-perceived ability	2.474	4	185	**.046**
	Geography enjoyment	.763	4	187	.551
	Geography classroom environment	1.403	4	189	.235
	Geography S-T relations	1.624	4	189	.170
	Geography peer competition	1.002	4	188	.408
	Geography classroom chaos	1.389	4	187	.239
	Geography homework behaviour	.811	4	188	.520
	Geography homework feedback	1.637	4	188	.167
	Geography homework total scale	1.268	4	188	.284
	Geography environment	.474	4	181	.755
	Geography usefulness	2.114	4	190	.081
	Geography anxiety	.514	4	187	.725
Time 3	Geography performance	.951	4	180	.436
	Geography self-perceived ability	.837	4	179	.503
	Geography enjoyment	2.224	4	173	.068
	Geography classroom environment	3.534	4	179	**.008**
	Geography S-T relations	1.580	4	179	.182
	Geography peer competition	2.749	4	178	**.030**
	Geography classroom chaos	2.792	4	177	.028
	Geography homework behaviour	.149	4	171	.963
	Geography homework feedback	1.937	4	171	.106
	Geography homework total scale	.909	4	170	.460
	Geography environment	.730	4	173	.572
	Geography usefulness	1.422	4	172	.229
	Geography anxiety	2.751	4	177	**.030**

Bold = significant at $p \leq .05$. **S-T** = student-teacher

Table 5.3.1. Maths classroom variables at time 1 (UK sample): Means, standard deviation (SD), and N with ANOVA results by classroom

Construct	C1	C2	C3	R1	R2	R3	*p*	η_p^2
Maths Primary school achievement	0.58 (0.91) n=25	0.78 (0.88) n=23	0.70 (0.88) n=20	0.54 (1.03) n=18	0.72 (1.09) n=21	0.81 (0.95) n=16	.942	.010
Maths school achievement	1.45 (0.66) n=25	0.36 (0.51) n=32	-0.52 (0.41) n=22	1.27 (0.49) n=24	0.12 (0.54) n=32	-0.26 (0.92) n=21	**.000**	**.586**
Maths performance	0.91 (0.60) n=26	0.24 (0.77) n=32	-0.58 (0.63) n=23	0.67 (0.58) n=24	-0.25 (0.80) n=32	-0.47 (0.96) n=21	**.000**	**.366**
Number line	-0.29 (0.71) n=26	-0.15 (0.82) n=32	0.24 (1.07) n=22	-0.37 (0.75) n=24	-0.17 (0.83) n=32	0.61 (0.82) n=20	**.001**	**.129**
Maths self-perceived ability	0.21 (0.93) n=26	0.01 (0.77) n=32	-0.08 (1.03) n=21	0.19 (0.68) n=24	-0.04 (1.00) n=31	0.21 (1.130 n=21	.757	.017
Maths enjoyment	0.07 (1.03) n=26	-0.13 (0.89) n=32	0.28 (0.98) n=22	0.23 (0.60) n=23	-0.03 (0.97) n=31	0.25 (1.02) n=21	.502	.028
Maths classroom environment	0.16 (0.71) n=26	0.20 (0.73) n=32	0.20 (0.89) n=23	0.13 (0.73) n=24	0.03 (1.01) n=32	0.42 (0.74) n=21	.696	.020
Maths classroom Student-teacher relations	0.11 (0.82) n=26	0.06 (0.81) n=32	0.21 (1.00) n=23	0.15 (0.92) n=24	-0.03 (0.95) n=32	0.40 (0.82) n=21	.660	.021

Significant results in bold following a Bonferroni multiple testing correction of $p \leq .000$ (p = .05 divided by number of measures (k=98) across maths and geography, time 1, 2 & 3)

Table 5.3.1. Continued. Maths classroom variables at time 1 (UK sample): Means, standard deviation (SD), and N with ANOVA results by classroom

Construct	**C1**	**C2**	**C3**	**R1**	**R2**	**R3**	***p***	**η_p^2**
Maths classroom peer competition	0.17 (0.98) n=26	0.31 (1.05) n=32	0.07 (0.68) n=23	0.00 (0.90) n=24	0.11 (0.98) n=32	0.21 (1.04) n=21	.875	.012
Maths classroom chaos	0.11 (0.88) n=26	-0.16 (0.90) n=32	-0.02 (0.78) n=23	-0.28 (1.12) n=24	-0.19 (0.96) n=32	-0.04 (1.10) n=21	.722	.018
Maths homework behaviour	-0.10 (1.07) n=26	0.26 (0.97) n=32	-0.03 (0.98) n=22	-0.10 (0.89) n=24	-0.34 (0.91) n=32	0.45 (0.88) n=21	.045	.072
Maths homework feedback	-0.28 (0.92) n=26	-0.57 (0.88) n=32	-0.12 (1.23) n=23	0.06 (0.67) n=24	-0.13 (0.83) n=32	0.48 (0.81) n=20	.003	.109
Maths homework total scale	-0.21 (1.04) n=26	-0.60 (0.99) n=32	-0.04 (1.23) n=23	0.12 (0.66) n=24	0.06 (0.87) n=32	0.21 (0.77) n=21	.022	.082
Maths environment	0.21 (0.79) n=26	-0.01 (0.93) n=32	-0.28 (1.02) n=22	0.43 (0.79) n=24	0.33 (1.00) n=32	0.30 (0.92) n=21	.085	.062
Maths usefulness	-0.19 (1.23) n=26	0.01 (0.76) n=31	0.04 (0.97) n=23	-0.08 (1.13) n=24	0.01 (1.16) n=31	0.11 (1.03) n=21	.949	.008
Maths anxiety	-0.35 (0.90) n=26	0.10 (0.83) n=32	0.57 (1.00) n=23	-0.40 (0.97) n=24	0.32 (1.19) n=31	0.43 (0.92) n=21	**.001**	**.122**

Significant results in bold following a Bonferroni multiple testing correction of $p \leq .000$ (p = .05 divided by number of measures (k=98) across maths and geography, time 1, 2 & 3)

Table 5.3.2. Maths classroom at time 1 (UK sample): Means, standard deviation (SD), for cognitive ability test, perceptions of intelligence and socioeconomic status and N with ANOVA results by maths classroom

Construct	C1	C2	C3	R1	R2	R3	*p*	η_p^2
Cognitive ability test	0.87 (0.57) n=23	-0.13 (0.53) n=24	-0.97 (0.60) n=15	0.87 (0.46) n=24	-0.02 (0.67) n=28	-0.94 (1.18) n=12	**.000**	**.537**
Theories of intelligence	0.72 (0.63) n=26	-0.16 (1.06) n=30	-0.16 (1.00) n=22	-0.03 (0.96) n=24	0.06 (0.95) n=31	-0.44 (0.89) n=20	**.001**	**.129**
Perceptions of academic and socio-economic mean score	0.09 (0.91) n=26	0.05 (0.75) n=31	0.22 (1.13) n=22	0.35 (0.73) n=23	-0.32 (1.02) n=32	-0.21 (1.35) n=19	.142	.054
Perceptions Of School Respect	-0.42 (0.97) n=26	-0.36 (0.77) n=29	0.32 (0.91) n=21	-0.01 (1.06) n=23	-0.50 (1.25) n=30	-0.15 (1.20) n=18	.076	.068
Perceptions Of School Grades	0.10 (1.02) n=26	-0.28 (0.86) n=28	-0.34 (0.69) n=20	0.57 (0.53) n=23	-0.18 (0.86) n=30	-0.38 (0.99) n=17	**.001**	**.131**
Perceptions of family occupation	-0.25 (1.00) n=23	-0.17 (0.86) n=28	0.16 (1.25) n=22	-0.06 (0.53) n=23	-0.35 (1.04) n=27	-0.05 (0.97) n=17	.549	.029
Perceptions of family education	0.12 (0.80) n=25	0.22 (0.78) n=30	0.06 (1.05) n=21	-0.19 (0.66) n=23	-0.32 (0.95) n=29	-0.10 (1.24) n=15	.225	.049

Significant results in bold following a Bonferroni multiple testing correction of $p \leq .000$ ($p = .05$ divided by number of measures (k=98) across maths and geography, time 1, 2 & 3)

Table 5.3.3. Geography classroom variables at time 1 (UK sample): Means, standard deviation (SD), and N with ANOVA results by classroom

Construct	7A	7C	7E	7H	7R	7T	*p*	η_p^2
Geography performance	-0.53 (0.95) n=23	-0.25 (0.85) n=28	0.09 (1.17) n=23	-0.60 (0.85) n=27	-0.32 (0.92) n=24	-0.28 (0.94) n=27	.159	.052
Geography self-perceived ability	0.38 (0.89) n=22	0.18 (1.03) n=24	-0.02 (1.09) n=18	0.08 (0.94) n=27	0.13 (1.13) n=23	0.19 (1.02) n=22	.980	.006
Geography enjoyment	0.08 (0.94) n=22	0.36 (0.92) n=24	0.06 (0.85) n=18	-0.22 (0.77) n=27	-0.12 (1.22) n=24	0.24 (0.95) n=24	.282	.045
Geography classroom environment	-0.08 (0.82) n=22	0.43 (0.65) n=23	0.32 (1.19) n=16	-0.08 (0.88) n=27	-0.45 (0.79) n=23	0.38 (0.93) n=17	.006	.123
Geography classroom Student-teacher relations	-0.34 (0.94) n=22	0.52 (0.77) n=23	0.19 (1.09) n=16	-0.17 (0.88) n=27	-0.61 (0.73) n=23	0.19 (0.92) n=17	**.000**	**.165**
Geography classroom peer competition	0.40 (0.95) n=22	0.09 (0.80) n=23	0.42 (1.20) n=16	0.13 (0.84) n=27	0.04 (0.88) n=23	0.53 (0.86) n=17	.414	.040

Significant results in bold following a Bonferroni multiple testing correction of $p \leq .000$ ($p = .05$ divided by number of measures (k=98) across maths and geography, time 1, 2 & 3)

Table 5.3.3. Continued. Geography classroom variables at time 1 (UK sample): Means, standard deviation (SD), and N with ANOVA results by classroom

Construct	7A	7C	7E	7H	7R	7T	*p*	η_p^2
Geography classroom chaos	-0.44 (0.91) n=22	0.65 (0.66) n=24	0.31 (0.98) n=16	0.08 (0.84) n=24	-0.16 (0.90) n=25	0.38 (0.70) n=16	**.000**	**.169**
Geography homework behaviour	-0.05 (1.12) n=22	0.07 (1.10) n=23	-0.06 (1.00) n=16	-0.01 (0.90) n=24	-0.17 (0.93) n=25	-0.19 (0.86) n=17	.958	.009
Geography homework feedback	0.02 (0.86) n=22	0.27 (0.95) n=24	0.82 (0.76) n=16	-0.17 (0.99) n=24	0.04 (0.58) n=24	0.28 (0.94) n=17	.015	.108
Geography homework total scale	0.01 (1.00) n=22	0.16 (1.06) n=24	0.71 (0.81) n=16	-0.12 (1.06) n=24	0.13 (0.65) n=24	0.35 (0.90) n=17	.120	.069
Geography environment	-0.40 (0.85) n=20	-0.11 (0.91) n=24	0.01 (0.86) n=16	-0.17 (0.85) n=24	-0.44 (0.81) n=24	-0.23 (0.93) n=16	.561	.032
Geography usefulness	-0.11 (1.12) n=22	0.20 (0.65) n=22	0.46 (0.96) n=16	-0.04 (0.91) n=24	-0.18 (0.90) n=23	0.28 (0.95) n=17	.224	.056
Geography anxiety	0.17 (0.78) n=22	-0.03 (0.87) n=24	-0.19 (0.99) n=21	-0.21 (0.94) n=23	0.29 (1.13) n=24	0.01 (0.90) n=23	.425	.036

Significant results in bold following a Bonferroni multiple testing correction of $p \leq .000$ ($p = .05$ divided by number of measures (k=98) across maths and geography, time 1, 2 & 3)

Table 5.3.4. Geography classroom at time 1 (UK sample): Means, standard deviation (SD), for cognitive ability test, perceptions of intelligence and socioeconomic status and N with ANOVA results by geography classroom

Construct	7A	7C	7E	7H	7R	7T	*p*	η_p^2
Cognitive ability test	0.26 (0.70) n=19	-0.09 (1.02) n=22	-0.20 (1.21) n=16	0.19 (1.02) n=26	0.48 (0.57) n=27	-0.21 (0.89) n=26	.054	.079
Theories of intelligence	0.17 (1.07) n=22	0.16 (0.99) n=28	-0.01 (1.02) n=23	0.09 (1.00) n=26	-0.12 (0.96) n=24	-0.27 (0.91) n=26	.561	.027
Perceptions of academic and socio-economic status mean score	-0.07 (0.97) n=22	0.28 (0.82) n=28	0.10 (1.01) n=23	0.05 (0.97) n=27	-0.09 (1.29) n=24	-0.20 (0.94) n=25	.585	.026
Perceptions Of School Respect	-0.58 (1.13) n=21	0.00 (0.77) n=28	-0.07 (1.12) n=21	-0.15 (0.84) n=25	-0.22 (1.25) n=24	-0.35 (1.18) n=25	.480	.032
Perceptions Of School Grades	-0.20 (0.91) n=20	0.00 (0.85) n=27	0.13 (0.83) n=22	-0.28 (1.00) n=25	-0.13 (1.01) n=23	0.01 (0.80) n=24	.665	.023
Perceptions of family occupation	-0.22 (1.03) n=22	-0.09 (0.97) n=26	-0.30 (1.13) n=21	0.06 (1.05) n=23	0.17 (0.74) n=20	-0.34 (0.66) n=25	.410	.037
Perceptions of family education	0.08 (0.98) n=22	0.24 (0.76) n=27	0.10 (0.89) n=19	0.03 (0.87) n=25	-0.25 (1.00) n=22	-0.47 (0.87) n=25	.066	.073

Significant results in bold following a Bonferroni multiple testing correction of $p \leq .000$ ($p = .05$ divided by number of measures (k=98) across maths and geography, time 1, 2 & 3)

Table 5.3.5. Maths classroom variables at time 1 (UK sample): Classrooms ranked by means (highest = 1 to lowest = 6) for measures demonstrating a significant effect of maths classroom

Construct	C1	C2	C3	R1	R2	R3	*p*	η_p^2
Cognitive ability test	**2nd**	**4th**	**6th**	**1st**	**3rd**	**5th**	**.000**	**.537**
Maths school achievement	**1st**	**3rd**	**6th**	**2nd**	**4th**	**5th**	**.000**	**.586**
Maths performance	**1st**	**3rd**	**6th**	**2nd**	**4th**	**5h**	**.000**	**.366**
Number line	**2nd**	**4th**	**5th**	**1st**	**3rd**	**6th**	**.001**	**.129**
Maths homework feedback*	**5th**	**6th**	**3rd**	**2nd**	**4th**	**1st**	.003	.109
Maths anxiety	**5th**	**4th**	**1st**	**6th**	**3rd**	**2nd**	**.001**	**.122**
Theories of intelligence	**1st**	**4th**	**5th**	**3rd**	**2nd**	**6th**	**.001**	**.129**
Perceptions Of school grades	**2nd**	**4th**	**5th**	**1st**	**3rd**	**6th**	**.001**	**.131**

Significant results in bold following a Bonferroni multiple testing correction of $p \leq .000$ (p = .05 divided by number of measures (k=98) across maths and geography, time 1, 2 & 3) *Just below significance but ranked for comparison. Anxiety: high score = high anxiety.

Table 5.3.6.Geography classroom variables at time 1 (UK sample): Classrooms ranked by means (highest = 1 to lowest = 6) for measures demonstrating a significant effect of geography classroom

Construct	7A	7C	7E	7H	7R	7T	*p*	η_p^2
Cognitive ability test**	2nd	4th	5th	3rd	1st	6th	.054	.079
Geography classroom environment*	4th	1st	3rd	5th	6th	2nd	.006	.123
Geography classroom Student-teacher relations	5th	1st	2nd	4th	6th	3rd	**.000**	**.165**
Geography classroom chaos	6th	1st	3rd	4th	5th	2nd	**.000**	**.169**

Significant results in bold following a Bonferroni multiple testing correction of $p \leq .000$ ($p = .05$ divided by number of measures (k=98) across maths and geography, time 1, 2 & 3) *Just below significance/**not significant but ranked for comparison. Chaos: high score = low chaos.

Table 5.3.7. Maths classroom variables at time 2 (UK sample): Means, standard deviation (SD), and N with ANOVA results by classroom

Construct	C1	C2	C3	R1	R2	R3	*p*	η_p^2
Maths school achievement	1.72 (0.62) n=27	0.33 (0.34) n=29	-0.78 (0.35) n=22	1.05 (0.40) n=30	0.31 (0.39) n=25	-0.60 (0.49) n=20	**.000**	**.795**
Maths performance	0.85 (0.59) n=27	0.00 (1.06) n=30	-0.72 (0.79) n=22	0.80 (0.55) n=30	-0.23 (0.86) n=25	-0.57 (0.84) n=21	**.000**	**.374**
Number line	-0.50 (0.75) n=27	-0.13 (0.86) n=28	0.06 (0.76) n=22	-0.31 (0.48) n=29	-0.13 (0.83) n=25	0.70 (0.87) n=21	**.000**	**.186**
Maths self-perceived ability	0.12 (1.03) n=27	0.05 (0.73) n=28	-0.06 (0.97) n=22	0.22 (0.75) n=28	0.08 (0.97) n=24	-0.03 (1.33) n=21	.923	.010
Maths enjoyment	-0.10 (0.93) n=27	-0.54 (1.00) n=30	0.22 (0.98) n=22	0.00 (0.91) n=28	-0.20 (0.82) n=25	0.03 (1.15) n=20	.100	.061
Maths classroom environment	0.15 (0.81) n=27	-0.10 (0.87) n=30	0.46 (0.91) n=22	0.17 (0.62) n=29	-0.09 (1.13) n=25	0.28 (0.92) n=21	.203	.047
Maths classroom Student-teacher relations	0.08 (0.77) n=27	-0.13 (0.81) n=30	0.36 (0.89) n=22	0.11 (0.73) n=29	-0.02 (1.05) n=25	0.33 (1.02) n=21	.318	.039
Maths classroom peer competition	0.16 (0.96) n=27	0.03 (1.06) n=30	0.33 (0.85) n=22	0.16 (0.97) n=29	-0.17 (1.00) n=25	0.01 (1.11) n=21	.622	.023

Significant results in bold following a Bonferroni multiple testing correction of $p \leq .000$ ($p = .05$ divided by number of measures (k=98) across maths and geography, time 1, 2 & 3)

Table 5.3.7. Continued. Maths classroom variables at time 2 (UK sample): Means, standard deviation (SD), and N with ANOVA results by classroom

Construct	C1	C2	C3	R1	R2	R3	*p*	η_p^2
Maths classroom chaos	-0.11 (0.96) n=27	-0.53 (0.90) n=29	-0.25 (0.87) n=22	-0.15 (1.03) n=29	-0.13 (0.92) n=24	0.35 (0.86) n=21	.055	.071
Maths homework behaviour	0.09 (1.20) n=27	0.10 (1.00) n=29	0.33 (0.80) n=22	0.06 (0.80) n=29	-0.27 (0.89) n=24	0.03 (1.00) n=21	.461	.031
Maths homework feedback	-0.35 (0.80) n=27	-0.24 (0.89) n=29	0.23 (1.08) n=22	-0.40 (0.63) n=29	0.23 (0.80) n=24	0.45 (1.00) n=21	**.001**	**.128**
Maths homework total scale	-0.37 (1.06) n=27	-0.24 (0.91) n=29	0.10 (1.07) n=22	-0.32 (0.59) n=29	0.35 (0.82) n=24	0.39 (1.12) n=21	.008	.101
Maths environment	0.26 (0.74) n=27	-0.06 (0.74) n=27	0.34 (1.03) n=22	0.46 (0.79) n=29	0.43 (1.10) n=24	0.16 (0.84) n=21	.271	.043
Maths usefulness	-0.25 (0.85) n=26	0.13 (0.93) n=26	0.28 (1.00) n=22	-0.09 (1.01) n=29	-0.04 (1.01) n=25	-0.09 (1.22) n=21	.538	.028
Maths anxiety	-0.54 (0.81) n=27	0.12 (0.96) n=28	0.51 (1.07) n=22	-0.30 (0.92) n=29	0.19 (1.11) n=25	0.37 (0.79) n=20	**.001**	**.133**

Significant results in bold following a Bonferroni multiple testing correction of $p \leq .000$ ($p = .05$ divided by number of measures (k=98) across maths and geography, time 1, 2 & 3)

Table 5.3.8. Geography classroom variables at time 2 (UK sample): Means, standard deviation (SD), and N with ANOVA results by classroom

Construct	7A	7C	7E	7H	7R	7T	*p*	η_p^2
Geography performance	-0.12 (1.01) n=25	-0.15 (1.22) n=25	0.07 (1.08) n=22	-0.21 (1.25) n=28	-0.11 (1.13) n=25	-0.26 (0.94) n=25	.948	.008
Geography self-perceivedability	-0.17 (0.99) n=23	0.09 (0.90) n=23	-0.14 (1.04) n=19	-0.30 (0.93) n=26	-0.11 (1.16) n=21	0.30 (0.82) n=24	.336	.042
Geography enjoyment	-0.03 (1.01) n=23	0.15 (0.95) n=23	-0.14 (0.90) n=21	-0.33 (0.84) n=26	-0.04 (1.11) n=23	0.20 (0.93) n=24	.420	.036
Geography classroom environment	0.06 (0.96) n=23	0.61 (0.87) n=23	0.01 (1.01) n=18	0.03 (0.69) n=28	0.20 (0.86) n=21	0.00 (0.80) n=23	.130	.063
Geography classroom Student-teacher relations	0.03 (0.97) n=23	0.65 (0.81) n=23	0.08 (1.06) n=18	-0.07 (0.82) n=28	0.16 (0.88) n=21	-0.06 (0.90) n=23	.075	.073
Geography classroom peer competition	0.11 (0.97) n=23	0.31 (0.96) n=23	-0.12 (0.91) n=18	0.17 (0.88) n=28	0.14 (0.93) n=21	0.09 (0.82) n=23	.796	.018

Significant results in bold following a Bonferroni multiple testing correction of $p \leq .000$ ($p = .05$ divided by number of measures (k=98) across maths and geography, time 1, 2 & 3)

Table 5.3.8. Continued. Geography classroom variables at time 2 (UK sample): Means, standard deviation (SD), and N with ANOVA results by classroom

Construct	**7A**	**7C**	**7E**	**7H**	**7R**	**7T**	***p***	**η_p^2**
Geography classroom chaos	-0.10 (0.84) n=21	0.54 (0.85) n=23	0.53 (0.76) n=17	0.00 (1.02) n=28	0.48 (0.87) n=22	0.32 (1.00) n=23	.064	.077
Geography homework behaviour	-0.39 (0.92) n=22	0.04 (0.99) n=22	-0.05 (1.00) n=17	0.16 (0.90) n=28	-0.23 (0.96) n=22	-0.18 (0.92) n=23	.383	.040
Geography homework feedback	0.15 (1.01) n=21	0.45 (1.08) n=22	0.41 (0.88) n=17	-0.12 (1.00) n=28	-0.02 (0.60) n=22	-0.13 (0.81) n=22	.130	.065
Geography homework total scale	0.32 (0.98) n=21	0.38 (1.03) n=22	0.38 (1.08) n=17	-0.16 (0.97) n=28	0.10 (0.70) n=22	-0.02 (0.87) n=22	.246	.051
Geography environment	-0.06 (0.91) n=21	0.15 (0.91) n=21	0.03 (0.96) n=17	0.00 (1.03) n=27	0.31 (0.78) n=22	0.17 (0.96) n=22	.812	.018
Geography usefulness	0.18 (0.98) n=21	-0.02 (0.81) n=21	0.11 (1.02) n=16	-0.10 (0.94) n=27	-0.02 (0.87) n=22	0.08 (0.70) n=22	.902	.013
Geography anxiety	-0.07 (0.76) n=21	0.11 (1.05) n=20	-0.23 (0.95) n=20	-0.28 (1.01) n=27	0.05 (1.10) n=22	0.15 (1.04) n=23	.587	.029

Significant results in bold following a Bonferroni multiple testing correction of $p \leq .000$ ($p = .05$ divided by number of measures (k=98) across maths and geography, time 1, 2 & 3)

Table 5.3.9. Maths classroom variables at time 2 (UK sample): Classrooms ranked by means (highest = 1 to lowest = 6) for measures demonstrating a significant effect of maths classroom

Construct	C1	C2	C3	R1	R2	R3	*p*	η_p^2
Maths school achievement	1st	3rd	6th	2nd	4th	5th	**.000**	**.795**
Maths performance	1st	3rd	6th	2nd	4th	5th	**.000**	**.374**
Number line	1st	4th	5th	2nd	3rd	6th	**.000**	**.186**
Maths homework feedback	5th	4th	3rd	6th	2nd	1st	**.001**	**.128**
Maths homework total scale*	6th	4th	3rd	5th	2nd	1st	.008	.101
Maths anxiety	6th	4th	1st	5th	3rd	2nd	**.001**	**.133**

Significant results in bold following a Bonferroni multiple testing correction of $p \leq .000$ ($p = .05$ divided by number of measures (k=98) across maths and geography, time 1, 2 & 3). Anxiety: high score = high anxiety. *Just below significance but ranked for comparison.

Table 5.3.10. Maths classroom variables at time 3 (UK sample): Means, standard deviation (SD), and N with ANOVA results by classroom

Construct	C1	C2	C3	C4	R1	R2	R3	R4	*p*	η_p^2
Maths performance	0.90 (0.69) n=28	0.28 (0.67) n=26	-0.30 (0.80) n=19	-1.12 (0.84) n=10	0.78 (0.56) n=26	-0.03 (0.66) n=24	-0.72 (0.83) n=18	-1.03 (0.77) n=13	**.000**	**.495**
Number line	-0.67 (0.61) n=27	-0.35 (0.71) n=26	0.18 (0.80) n=19	0.31 (0.82) n=10	-0.53 (0.70) n=26	0.07 (0.72) n=23	0.09 (0.98) n=18	0.67 (0.64) n=13	**.000**	**.242**
Maths self-perceived ability	0.22 (0.85) n=25	0.02 (0.78) n=25	-0.21 (0.83) n=17	-0.22 (0.95) n=10	0.02 (0.73) n=26	-0.03 (0.93) n=24	-0.04 (1.05) n=18	-0.27 (0.96) n=12	.738	.028
Maths enjoyment	-0.04 (0.78) n=26	-0.51 (1.19) n=26	-0.13 (0.99) n=18	0.12 (1.05) n=10	0.05 (0.72) n=26	-0.21 (0.86) n=24	-0.10 (0.93) n=17	-0.20 (0.96) n=11	.485	.042
Maths classroom environment	0.58 (0.83) n=27	-0.14 (1.00) n=26	0.42 (1.02) n=18	-0.25 (0.98) n=10	0.25 (0.76) n=26	0.48 (0.78) n=24	-0.35 (0.67) n=18	-0.26 (0.71) n=12	**.001**	**.149**
Maths classroom Student-teacher relations	0.68 (0.85) n=27	-0.22 (0.99) n=26	0.39 (0.99) n=18	-0.09 (1.17) n=10	0.25 (0.83) n=26	0.37 (0.70) n=24	-0.18 (0.64) n=18	-0.04 (1.04) n=12	.006	.120
Maths classroom peer competition	0.11 (0.75) n=27	0.09 (1.03) n=26	0.27 (0.90) n=18	-0.41 (0.64) n=10	0.12 (0.92) n=26	0.45 (0.86) n=24	-0.47 (0.64) n=18	-0.51 (0.89) n=12	.005	.121

Significant results in bold following a Bonferroni multiple testing correction of $p \leq .000$ ($p = .05$ divided by number of measures (k=98) across maths and geography, time 1, 2 & 3)

Table 5.3.10. Continued. Maths classroom variables at time 3 (UK sample): Means, standard deviation (SD), and N with ANOVA results by classroom

Construct	C1	C2	C3	C4	R1	R2	R3	R4	*p*	η_p^2
Maths classroom chaos	0.06 (0.99) n=27	-0.44 (1.01) n=26	-0.06 (0.87) n=19	0.11 (0.69) n=9	0.18 (0.92) n=25	0.47 (0.65) n=23	0.36 (0.55) n=18	0.05 (0.98) n=13	.025	.098
Maths homework behaviour	-0.51 (0.90) n=27	-0.09 (1.01) n=26	0.42 (0.89) n=19	0.56 (1.09) n=9	-0.30 (0.85) n=25	-0.08 (1.14) n=23	-0.11 (1.01) n=18	0.25 (0.93) n=13	.022	.100
Maths homework feedback	-0.25 (0.78) n=26	-0.16 (1.09) n=26	0.80 (0.84) n=19	0.20 (1.20) n=8	-0.45 (0.71) n=25	0.16 (0.92) n=23	-0.07 (0.87) n=18	0.58 (1.15) n=13	**.000**	**.164**
Maths homework total scale	-0.01 (0.84) n=26	-0.10 (1.13) n=26	0.49 (0.81) n=19	-0.19 (1.26) n=8	-0.26 (0.74) n=25	0.09 (1.03) n=23	-0.03 (0.72) n=18	0.40 (1.04) n=13	.184	.064
Maths environment	0.52 (0.70) n=27	-0.28 (1.08) n=26	0.27 (0.88) n=19	-0.11 (0.95) n=9	0.11 (0.66) n=25	0.55 (0.97) n=23	-0.24 (0.70) n=18	0.42 (0.94) n=12	.004	.128
Maths usefulness	0.18 (0.88) n=26	-0.12 (1.15) n=26	-0.07 (0.72) n=18	-0.51 (1.00) n=10	0.05 (0.94) n=26	-0.03 (1.00) n=24	-0.29 (0.67) n=18	-0.42 (1.01) n=12	.429	.044
Maths anxiety	-0.46 (0.76) n=27	-0.04 (0.94) n=26	0.26 (1.31) n=18	0.45 (0.65) n=10	-0.29 (0.75) n=26	0.01 (0.89) n=24	0.69 (1.23) n=18	0.27 (0.78) n=12	.003	.131

Significant results in bold following a Bonferroni multiple testing correction of $p \leq .000$ (p = .05 divided by number of measures (k=98) across maths and geography, time 1, 2 & 3)

Table 5.3.11. Geography classroom variables at time 3 (UK sample): Means, standard deviation (SD), and N with ANOVA results by classroom

Construct	7A	7C	7E	7H	7R	7T	*p*	η_p^2
Geography performance	0.08 (0.98) n=25	-0.42 (0.90) n=28	0.07 (1.17) n=25	-0.65 (1.06) n=28	-0.35 (1.06) n=24	-0.06 (0.97) n=27	.055	.069
Geography self-perceived ability	-0.26 (1.06) n=24	0.04 (0.98) n=28	-0.10 (1.12) n=24	0.08 (0.81) n=26	-0.01 (1.18) n=25	0.04 (0.95) n=27	.855	.013
Geography enjoyment	-0.13 (1.06) n=24	-0.03 (1.09) n=28	-0.11 (1.01) n=24	-0.04 (0.87) n=26	-0.05 (1.10) n=25	0.13 (1.04) n=27	.960	.007
Geography classroom environment	-0.07 (1.00) n=24	-0.08 (1.12) n=27	-0.02 (0.96) n=24	0.10 (0.91) n=27	-0.26 (0.95) n=24	0.02 (0.93) n=27	.863	.013
Geography classroom Student-teacher relations	-0.13 (0.96) n=24	-0.06 (1.10) n=27	0.03 (0.91) n=24	0.01 (0.94) n=27	-0.31 (0.94) n=24	-0.12 (0.88) n=27	.840	.014
Geography classroom peer competition	0.08 (0.99) n=24	-0.09 (1.08) n=27	-0.11 (1.03) n=24	0.20 (0.95) n=27	-0.07 (0.98) n=24	0.27 (0.98) n=27	.627	.023

Significant results in bold following a Bonferroni multiple testing correction of $p \leq .000$ ($p = .05$ divided by number of measures (k=98) across maths and geography, time 1, 2 & 3)

Table 5.3.11. Continued. Geography classroom variables at time 3 (UK sample): Means, standard deviation (SD), and N with ANOVA results by classroom

Construct	**7A**	**7C**	**7E**	**7H**	**7R**	**7T**	***p***	**η_p^2**
Geography classroom chaos	-0.17 (0.96) n=24	0.37 (0.88) n=27	0.32 (0.69) n=24	-0.24 (1.10) n=27	-0.29 (1.06) n=25	0.15 (1.02) n=27	.042	.074
Geography homework behaviour	0.07 (1.24) n=24	-0.19 (0.99) n=27	-0.17 (0.90) n=24	0.03 (0.99) n=27	0.09 (1.00) n=24	-0.29 (1.02) n=27	.689	.020
Geography homework feedback	-0.20 (0.99) n=24	-0.07 (0.96) n=27	0.23 (0.81) n=24	-0.35 (1.10) n=27	-0.54 (0.65) n=24	-0.16 (1.07) n=27	.105	.060
Geography homework total scale	-0.25 (1.19) n=24	0.02 (0.87) n=27	0.29 (0.62) n=24	-0.30 (1.04) n=27	-0.49 (0.73) n=24	-0.04 (1.08) n=27	.071	.066
Geography environment	0.16 (0.89) n=23	0.11 (1.13) n=26	-0.34 (0.93) n=24	-0.07 (1.14) n=27	0.10 (1.05) n=25	0.11 (0.93) n=27	.525	.028
Geography usefulness	0.12 (1.13) n=23	-0.27 (1.10) n=27	0.10 (0.70) n=23	-0.36 (0.92) n=27	-0.41 (1.06) n=25	0.08 (0.94) n=27	.162	.052
Geography anxiety	0.04 (0.95) n=24	0.14 (0.90) n=27	-0.24 (0.86) n=24	-0.14 (1.08) n=27	0.12 (1.18) n=25	0.09 (1.07) n=27	.716	.019

Significant results in bold following a Bonferroni multiple testing correction of $p \leq .000$ ($p = .05$ divided by number of measures (k=98) across maths and geography, time 1, 2 & 3)

Table 5.3.12. Maths classroom variables at time 3 (UK sample): Classrooms ranked by means (highest = 1 to lowest = 8) for measures demonstrating a significant effect of maths classroom

Construct	C1	C2	C3	C4	R1	R2	R3	R4	*p*	η_p^2
Maths performance	1st	3rd	5th	8th	2nd	4th	6th	7th	**.000**	**.495**
Number line	1st	3rd	6th	7th	2nd	4th	5th	8th	**.000**	**.242**
Maths classroom environment	1st	5th	3rd	6th	4th	2nd	8th	7th	**.001**	**.149**
Maths classroom Student-teacher relations*	1st	8th	2nd	6th	4th	3rd	7th	5th	.006	.120
Maths classroom peer competition*	4th	5th	2nd	6th	3rd	1st	7th	8th	.005	.121
Maths homework feedback	7th	6th	1st	3rd	8th	4th	5th	2nd	**.000**	**.164**
Maths Environment*	2nd	8th	4th	6th	5th	1st	7th	3rd	.004	.128
Maths anxiety*	8th	6th	4th	2nd	7th	5th	1st	3rd	.003	.131

Significant results in bold following a Bonferroni multiple testing correction of $p \leq .000$ ($p = .05$ divided by number of measures (k=98) across maths and geography, time 1, 2 & 3) *Just below significance but ranked for comparison. Anxiety: high score = high anxiety

Table 5.3.13. Levene's test of equality of variances for maths classroom measures without controlling for prior achievement at time 1 and time 2 (UK sample)

Wave	Construct	F	df1	df2	Sig.
Time 1	Cognitive ability by maths class	1.690	5	120	.142
	Maths Primary achievement	.433	5	117	.825
	Maths school achievement	2.020	5	150	.079
	Maths performance	1.337	5	152	.252
	Number line	1.074	5	150	.377
	Maths self-perceived ability	3.455	5	149	**.006**
	Maths enjoyment	1.434	5	149	.215
	Maths classroom environment	.831	5	152	.530
	Maths student-teacher relations	.702	5	152	.623
	Maths peer competition	1.290	5	152	.271
	Maths classroom chaos	1.160	5	152	.331
	Maths homework behaviour	.799	5	151	.552
	Maths homework feedback	2.839	5	151	**.018**
	Maths homework total scale	2.701	5	152	**.023**
	Maths environment	1.159	5	151	.332
	Maths usefulness	1.283	5	150	.274
	Maths anxiety	1.517	5	151	.188
	Theories of intelligence	2.619	5	147	**.027**
	Perceptions of academic and socioeconomic status	2.716	5	147	**.022**
	Self-perceptions of school respect	2.313	5	141	.047
	Self-perceptions of school grades	1.586	5	138	.168
	Self-perceptions of family SES, occupation	2.110	5	134	.068
	Self-perceptions of family SES, education	3.333	5	137	**.007**
Time 2	Maths school achievement	4.354	5	147	**.001**
	Maths performance	2.503	5	149	**.033**
	Number line	.930	5	146	.464
	Maths self-perceived ability	3.464	5	144	**.005**
	Maths enjoyment	.754	5	146	.585
	Maths classroom environment	1.489	5	148	.197
	Maths student-teacher relations	1.775	5	148	.121
	Maths peer competition	.523	5	148	.759
	Maths classroom chaos	.140	5	146	.983
	Maths homework behaviour	2.341	5	146	**.044**
	Maths homework feedback	1.819	5	146	.113
	Maths homework total scale	2.400	5	146	**.040**
	Maths environment	2.386	5	144	**.041**
	Maths usefulness	.611	5	143	.691
	Maths anxiety	1.099	5	145	.364

Bold = significant at $p \leq .05$

Table 5.3.14. Levene's test of equality of variances for geography classroom measures without controlling for prior achievement at time 1 and time 2 (UK sample)

Wave	Construct	F	df1	df2	Sig.
Time 1	Cognitive ability geography class	4.448	5	130	.001
	Geography performance	1.152	5	146	.336
	Geography self-perceived ability	.926	5	130	.467
	Geography enjoyment	2.126	5	133	.066
	Geography classroom environment	.886	5	122	.493
	Geography S-T relations	.883	5	122	.495
	Geography peer competition	1.044	5	122	.395
	Geography classroom chaos	.749	5	121	.589
	Geography homework behaviour	.953	5	121	.449
	Geography homework feedback	1.387	5	121	.234
	Geography homework total scale	1.135	5	121	.345
	Geography environment	.328	5	118	.895
	Geography usefulness	1.133	5	118	.347
	Geography anxiety	1.146	5	131	.340
	Theories of intelligence	.353	5	143	.880
	Perceptions of academic and socioeconomic status	2.101	5	143	.069
	Self-perceptions of school respect	3.020	5	138	**.013**
	Self-perceptions of school grades	.662	5	135	.653
	Self- perceptions of family SES, occupation	1.765	5	131	.124
	Self- perceptions of family SES, education	.283	5	134	.922
Time 2	Geography performance	.587	5	144	.710
	Geography self-perceived ability	.728	5	130	.604
	Geography enjoyment	.566	5	134	.726
	Geography classroom environment	.613	5	130	.690
	Geography S-T relations	.440	5	130	.820
	Geography peer competition	.214	5	130	.956
	Geography classroom chaos	.717	5	128	.612
	Geography homework behaviour	.198	5	128	.963
	Geography homework feedback	2.006	5	126	.082
	Geography homework total scale	1.004	5	126	.418
	Geography environment	.452	5	124	.811
	Geography usefulness	.871	5	123	.503
	Geography anxiety	1.349	5	127	.248

Bold = significant at $p \leq .05$. **S-T** = student-teacher

Table 5.3.15. Levene's test of equality of variances for maths classroom measures without controlling for prior achievement at time 3 (UK sample)

Construct	F	df1	df2	Sig.
Maths performance	.638	7	156	.724
Number line	.780	7	154	.605
Maths self-perceived ability	.634	7	149	.727
Maths enjoyment	1.277	7	150	.266
Maths classroom environment	.562	7	153	.786
Maths student-teacher relations	1.959	7	153	.064
Maths peer competition	1.109	7	153	.360
Maths classroom chaos	1.103	7	152	.364
Maths homework behaviour	.680	7	152	.689
Maths homework feedback	1.422	7	150	.201
Maths homework total scale	1.595	7	150	.141
Maths environment	1.203	7	151	.305
Maths usefulness	1.353	7	152	.230
Maths anxiety	3.029	7	153	**.005**

Bold = significant at $p \leq .05$

Table 5.3.16. Levene's test of equality of variances for geography classroom measures without controlling for prior achievement at time 3 (UK sample)

Construct	F	df1	df2	Sig.
Geography performance	.883	5	151	.494
Geography self-perceived ability	.890	5	148	.490
Geography enjoyment	.503	5	148	.773
Geography classroom environment	.378	5	147	.863
Geography student-teacher relations	.442	5	147	.819
Geography peer competition	.283	5	147	.922
Geography classroom chaos	1.506	5	148	.191
Geography homework behaviour	2.164	5	147	.061
Geography homework feedback	1.519	5	147	.187
Geography homework total scale	2.375	5	147	**.042**
Geography environment	.805	5	146	.548
Geography usefulness	.963	5	146	.443
Geography anxiety	.643	5	148	.667

Bold = significant at $p \leq .05$

Table 5.4.1. Means and standard deviation (SD) and N for teacher characteristics

	Teacher's age at time of testing	Years of teaching experience	Teacher self efficacy in student engagement	Teacher self efficacy in instructional strategies	Teacher self efficacy in classroom management	Emotional ability
N	14	17	17	17	17	17
Mean	49.93	25.00	6.53	7.49	6.98	5.27
SD	7.87	8.69	1.29	0.78	1.14	0.30
Minimum	35	12	4	6	5	4.77
Maximum	63	40	9	9	9	5.80

Table 5.4.2. Primary school teacher characteristics: Bivariate correlations between primary school teacher characteristics and measures showing a significant effect of **maths classroom at time 1** (N)

Construct	1	2	3	4	5	6	7	8
1. Maths primary school achievement	1 (219)							
2. Maths performance	**.310** (207)**	1 (229)						
3. Maths classroom student-teacher relations	**.155* (202)**	**.155* (223)**	1 (223)					
4. Maths classroom chaos	.035 (205)	**.182** (227)**	**.197** (221)**	1 (227)				
5. Years of teaching experience	-.005 (162)	.103 (178)	.085 (174)	**.204** (176)**	1 (192)			
6. Teacher emotional ability	**-.216** (183)**	**.172* (187)**	-.017 (183)	.048 (185)	-.012 (192)	1 (216)		
7. Teacher self-efficacy in student engagement	-.051 (183)	.088 (187)	-.003 (183)	**.156* (185)**	.139 (192)	**.429** (216)**	1 (216)	
8. Teacher self-efficacy in instructional strategies	-.131 (171)	.083 (187)	-.021 (183)	**.274** (185)**	**.529** (178)**	.094 (187)	**.516** (187)**	1 (187)
9. Teacher self-efficacy in classroom management	-.002 (183)	.072 (187)	-.073 (183)	.073 (185)	**-.208** (192)**	**.363** (216)**	**.923** (216)**	**.323** (187)**

Scale: Maths primary school achievement 1-5; **Maths performance** 0-48; **Student-teacher relations:** 0-3; **Classroom chaos:** 0-1;**Teaching experience:** 12-40 yrs; **Teacher emotional ability:**1-7;**Self-efficacy factors:** 1-9. **. Correlation is significant at the 0.01 level (2-tailed).*. Correlation is significant at the 0.05 level (2-tailed).

Table 5.4.3. Primary school teacher characteristics: Bivariate correlations between primary school teacher characteristics and measures showing a significant effect of **maths classroom at time 2** (N)

Construct	1	2	3	4	5	6	7	8	9	10
1. Maths year 5 school	1									
achievement	(225)									
2. Maths performance	**.409****	1								
	(201)	(222)								
3. Number line	**-.179***	**-.239****	1							
	(194)	**(200)**	(220)							
4. Maths classroom environment	.078	.084	-.042	1						
	(202)	(220)	(218)	(222)						
5. Maths classroom chaos	.116	**.259****	-.083	**.171***	1					
	(203)	**(222)**	(220)	**(222)**	(224)					
6. Maths homework feedback	.115	**.154***	.036	**.409****	**.197****	1				
	(199)	**(218)**	(216)	**(218)**	**(220)**	(220)				
7. Years of teaching experience	.142	**.170***	**-.264****	-.030	.065	-.002	1			
	(167)	**(168)**	**(165)**	(169)	(169)	(166)	(192)			
8. Teacher emotional ability	.034	.050	**-.157***	-.04	-.014	**-.263****	-.012	1		
	(188)	(188)	**(185)**	(189)	(189)	**(185)**	(192)	(216)		
9. Teacher self-efficacy in	.115	-.026	-.044	-.090	-.101	.010	.139	**.429****	1	
student engagement	(188)	(188)	(185)	(189)	(189)	(185)	(192)	**(216)**	(216)	
10. Teacher self-efficacy in	.146	.097	**-.284****	.055	**.220****	.114	**.529****	.094	**.516****	1
instructional strategies	(176)	(170)	**(167)**	(171)	**(171)**	(167)	**(178)**	(187)	**(187)**	(187)
11. Teacher self-efficacy in	.088	-.054	.013	-.128	**-.147***	.010	**-.208****	**.363****	**.923****	**.323****
classroom management	(188)	(188)	(185)	(189)	**(189)**	(185)	**(192)**	**(216)**	**(216)**	**(187)**

Scale: Maths year 5 school achievement 1-5; **Maths performance** 0-48; **Number line:** low score =optimum; **Classroom environment:** 0-3; **Classroom chaos:** 0-1; **Homework feedback:** 0-3; **Teaching experience:** 12-40 yrs; **Teacher emotional ability:** 1-7; **Self-efficacy factors:** 1-9. **. Correlation is significant at the 0.01 level (2-tailed). *. Correlation is significant at the 0.05 level (2-tailed).

Table 5.4.4. Primary school teacher characteristics: Bivariate correlations between primary school teacher characteristics and measures showing a significant effect of **maths classroom at time 3** (N)

Construct	1	2	3	4	5	6	7	8
1. Maths classroom environment	1 (218)							
2. Maths classroom student-teacher relations	**.941**** **(218)**	1 (218)						
3. Maths homework behaviour	**-.180**** **(215)**	**-.149*** **(215)**	1 (224)					
4. Maths classroom chaos	.096 (218)	.121 (218)	**-.216**** **(224)**	1 (227)				
5.Years of teaching experience	.000 (163)	.011 (163)	-.026 (167)	.084 (169)	1 (192)			
6. Teacher emotional ability	.117 (185)	.105 (185)	.040 (188)	.005 (191)	-.012 (192)	1 (216)		
7. Teacher self-efficacy in student engagement	-.024 (185)	-.036 (185)	.063 (188)	.020 (191)	.139 (192)	**.429**** **216)**	1 (216)	
8. Teacher self-efficacy in instructional strategies	.062 (157)	.041 (157)	-.020 (160)	**.252**** **(162)**	**.529**** **(178)**	.094 (187)	**.516**** **(187)**	1 (187)
9. Teacher self-efficacy in classroom management	-.081 (185)	-.098 (185)	.026 (188)	-.032 (191)	**-.208**** **(192)**	**.363**** **(216)**	**.923**** **(216)**	**.323**** **(187)**

Scale: Classroom environment: 0-3; **Student-teacher relations:** 0-3; **Homework behaviour:** 0-3; **Classroom chaos:** 0-1;**Teaching experience:** 12-40 yrs; **Teacher emotional ability:**1-7;**Self-efficacy factors:** 1-9. **. Correlation is significant at the 0.01 level (2-tailed).*. Correlation is significant at the 0.05 level (2-tailed).

Table 5.4.5. Primary school teacher characteristics: Bivariate correlations between primary school teacher characteristics and measures showing a significant effect of **geography classroom at time 1** (N)

Construct	1	2	3	4	5	6	7	8
1. Geography primary school achievement	1 (220)							
2. Geography classroom environment	**.171*** **(193)**	1 (214)						
3. Geography classroom student-teacher relations	**.188**** **(193)**	**.912**** **(214)**	1 (214)					
4. Geography classroom chaos	**.198**** **(197)**	**.190**** **(214)**	**.187**** **(214)**	1 (218)				
5. Years of teaching experience	**.207**** **(163)**	**.287**** **(165)**	**.305**** **(165)**	**.165*** **(168)**	1 (192)			
6. Teacher emotional ability	-.142 (184)	-.030 (174)	-.043 (174)	.026 (177)	-.012 (192)	1 (216)		
7. Teacher self-efficacy in student engagement	**.153*** **(184)**	.117 (174)	.135 (174)	.007 (177)	.139 (192)	**.429**** **(216)**	1 (216)	
8. Teacher self-efficacy in instructional strategies	.040 (172)	**.203**** **(174)**	**.183*** **(174)**	**.158*** **(177)**	**.529**** **(178)**	.094 (187)	**.516**** **(187)**	1 (187)
9. Teacher self-efficacy in classroom management	.143 (184)	.018 (174	.029 (174)	-.074 (177)	**-.208**** **(192)**	**.363**** **(216)**	**.923**** **(216)**	**.323**** **(187)**

Scale: Geography primary school achievement 1-5; **Classroom environment:** 0-3; **Student-teacher relations:** 0-3; **Classroom chaos:** 0-1;**Teaching experience:** 12-40 yrs; **Teacher emotional ability:**1-7;**Self-efficacy factors:** 1-9. **. Correlation is significant at the 0.01 level (2-tailed).*. Correlation is significant at the 0.05 level (2-tailed).

Table 5.4.6. Primary school teacher characteristics: Bivariate correlations between primary school teacher characteristics and measures showing a significant effect of **geography classroom at time 2** (N)

Construct	1	2	3	4	5	6	7
1. Geography performance	1 (218)						
2. Geography classroom student-teacher relations	.126 (213)	1 (219)					
3. Geography environment	**.228**** **(204)**	**.352**** **(209)**	1 (210)				
4. Years of teaching experience	.112 (165)	.103 (165)	**.253**** **(159)**	1 (192)			
5. Teacher emotional ability	**-.233**** **(185)**	.081 (185)	.050 (177)	-.012 (192)	1 (216)		
6. Teacher self-efficacy in student engagement	**-.238**** **(185)**	-.088 (185)	.028 (177)	.139 (192)	**.429**** **(216)**	1 (216)	
7. Teacher self-efficacy in instructional strategies	.115 (168)	.067 (167)	**.215**** **(159)**	**.529**** **(178)**	.094 (187)	**.516**** **(187)**	1 (187)
8. Teacher self-efficacy in classroom management	**-.235**** **(185)**	-.149* (185)	-.053 (177)	**-.208**** **(192)**	**.363**** **(216)**	**.923**** **(216)**	**.323**** **(187)**

Scale: Geography performance 0-37; **Student-teacher relations:** 0-3; **Teaching experience:** 12-40 yrs; **Teacher emotional ability:**1-7;**Self-efficacy factors:** 1-9.
**. Correlation is significant at the 0.01 level (2-tailed).*. Correlation is significant at the 0.05 level (2-tailed).

Table 5.4.7. Primary school teacher characteristics: Bivariate correlations between primary school teacher characteristics and measures showing a significant effect of **geography classroom at time 3** (N)

Construct	1	2	3	4	5	6	7	8
1. Geography performance	1 (224)							
2. Geography classroom environment	.107 (217)	1 (220)						
3. Geography classroom student-teacher relations	**.132*** **(220)**	**.905**** **(219)**	1 (223)					
4. Geography environment	**.162*** **(214)**	**.423**** **(212)**	**.382**** **(216)**	1 (216)				
5. Years of teaching experience	.084 (167)	.152 (163)	.121 (165)	**.253**** **(161)**	1 (192)			
6. Teacher emotional ability	-.138 (189)	-.075 (185)	.009 (187)	**.179*** **(183)**	-.012 (192)	1 (216)		
7. Teacher self-efficacy in student engagement	-.180* (189)	-.077 (185)	-.076 (187)	**.178*** **(183)**	.139 (192)	**.429**** **(216)**	1 (216)	
8. Teacher self-efficacy in instructional strategies	.019 (160)	**.165*** **(157)**	.136 (158)	**.290**** **(154)**	**.529**** **(178)**	.094 (187)	**.516**** **(187)**	1 (187)
9. Teacher self-efficacy in classroom management	**-.185*** **(189)**	-.155* (185)	**-.155*** **(187)**	.074 (183)	**-.208**** **(192)**	**.363**** **(216)**	**.923**** **(216)**	**.323**** **(187)**

Scale: Geography performance: 0-37; **Classroom environment:** 0-3; **Student-teacher relations**: 0-3; **Geography environment:** 1-4; **Geography anxiety:** 1-5; **Teaching experience:** 12-40 yrs; **Teacher emotional ability:** 1-7; **Self-efficacy factors:** 1-9. **. Correlation is significant at the 0.01 level (2-tailed).
*. Correlation is significant at the 0.05 level (2-tailed).

Table 5.4.8. Maths teacher characteristics: Bivariate correlations between maths teacher characteristics and measures showing a significant effect of **maths classroom at time 1** (N)

Construct	1	2	3	4	5	6	7	8
1. Maths primary school achievement	1 (219)							
2. Maths performance	**.310**** **(207)**	1 (229)						
3. Maths classroom student-teacher relations	.155* (202)	**.155*** **(223)**	1 (223)					
4. Maths classroom chaos	.035 (205)	**.182**** **(227)**	**.197**** **(221)**	1 (227)				
5. Years of teaching experience	.056 (200)	-.069 (204)	-.136 (201)	-.107 (202)	1 (235)			
6. Teacher emotional ability	**-.181*** **(200)**	-.056 (204)	**-.298**** **(201)**	-.089 (202)	.052 (235)	1 (235)		
7. Teacher self-efficacy in student engagement	**-.141*** **(200)**	**-.243**** **(204)**	.049 (201)	**-.142*** **(202)**	**.571**** **(235)**	-.163* (235)	1 (235)	
8. Teacher self-efficacy instructional strategies	.017 (200)	**-.236**** **(204)**	-.014 (201)	**-.166*** **(202)**	**.824**** **(235)**	**-.263**** **(235)**	**.837**** **(235)**	1 (235)
9. Teacher self-efficacy in classroom management	**-.193**** **(200)**	-.090 (204)	.094 (201)	-.040 (202)	**.400**** **(235)**	-.073 (235)	**.882**** **(235)**	**.541**** **(235)**

Scale: Maths primary school achievement 1-5; **Maths performance** 0-48; **Student-teacher relations:** 0-3; **Classroom chaos:** 0-1;**Teaching experience:** 12-40 yrs; **Teacher emotional ability:**1-7;**Self-efficacy factors:** 1-9. **. Correlation is significant at the 0.01 level (2-tailed).*. Correlation is significant at the 0.05 level (2-tailed).

Table 5.4.9. Maths teacher characteristics: Bivariate correlations between maths teacher characteristics and measures showing a significant effect of **maths classroom at time 2** (N)

Construct	1	2	3	4	5	6	7	8	9	10
1. Maths achievement year 5	1									
	(225)									
2. Maths performance	**.409****	1								
	(201)	(222)								
3. Number line	**-.239****	**-.315****	1							
	(200)	**(218)**	(220)							
4. Maths classroom environment	.078	.084	-.042	1						
	(202)	(220)	(218)	(222)						
5. Maths classroom chaos	.116	**.259****	-.083	**.171***	1					
	(203)	**(222)**	(220)	**(222)**	(224)					
6. Maths homework feedback	.115	**.154***	.036	**.409****	**.197****	1				
	(199)	**(218)**	(216)	**(218)**	**(220)**	(220)				
7. Years of teaching experience	-.004	-.123	.133	-.074	**-.159***	.109	1			
	(206)	(201)	(198)	(201)	**(202)**	(198)	(235)			
8. Teacher emotional ability	-.130	**-.146***	-.097	**-.200****	-.019	.025	.052	1		
	(206)	**(201)**	(198)	**(201)**	(202)	(198)	(235)	(235)		
9. Teacher self-efficacy in	-.104	**-.223****	**.317****	**.180***	-.066	.038	**.571****	-.163*	1	
student engagement	(206)	**(201)**	**(198)**	**(201)**	(202)	(198)	**(235)**	(235)	(235)	
10. Teacher self-efficacy in	-.073	**-.191****	**.276****	.087	**-.154***	.118	**.824****	**-.263****	**.837****	1
instructional strategies	(206)	**(201)**	**(198)**	(201)	**(202)**	(198)	**(235)**	**(235)**	**(235)**	(235)
11. Teacher self-efficacy in	-.040	**-.144***	**.243****	**.184****	.019	-.048	**.400****	-.073	**.882****	**.541****
classroom management	(206)	**(201)**	**(198)**	**(201)**	(202)	(198)	**(235)**	(235)	**(235)**	**(235)**

Scale: Maths school achievement 1-5; **Maths performance** 0-48; **Number line:** low score =optimum; **Classroom environment:** 0-3; **Classroom chaos:** 0-1;**Homework feedback:** 0-3; **Teaching experience:** 12-40 yrs; **Teacher emotional ability:**1-7; **Self-efficacy factors:** 1-9. **. Correlation is significant at the 0.01 level (2-tailed).*. Correlation is significant at the 0.05 level (2-tailed).

Table 5.4.10. Maths teacher characteristics: Bivariate correlations between maths teacher characteristics and measures showing a significant effect of **maths classroom at time 3** (N)

Construct	1	2	3	4	5	6	7	8
1. Maths classroom environment	1 (218)							
2. Maths classroom student-teacher	**.941**** **(218)**	1 (218)						
3. Maths homework behaviour	**-.180**** **(215)**	**-.149*** **(215)**	1 (224)					
4. Maths classroom chaos	.096 (218)	.121 (218)	**-.216**** **(224)**	1 (227)				
5. Years of teaching experience	**-.149*** **(199)**	**-.147*** **(199)**	-.018 (202)	-.049 (205)	1 (235)			
6. Teacher emotional ability	**-.156*** **(199)**	**-.153*** **(199)**	.107 (202)	-.026 (205)	.052 (235)	1 (235)		
7. Teacher self-efficacy student engagement	.101 (199)	.111 (199)	.058 (202)	-.014 (205)	**.571**** **(235)**	**-.163*** **(235)**	1 (235)	
8. Teacher self-efficacy in instructional strategies	-.047 (199)	-.053 (199)	.019 (202)	-.077 (205)	**.824**** **(235)**	**-.263**** **(235)**	**.837**** **(235)**	1 (235)
9. Teacher self-efficacy in classroom management	**.192**** **(199)**	**.218**** **(199)**	.037 (202)	.063 (205)	**.400**** **(235)**	-.073 (235)	**.882**** **(235)**	**.541**** **(235)**

Scale: Classroom environment: 0-3; **Student-teacher relations:** 0-3; **Classroom chaos:** 0-1;**Teaching experience:** 12-40 yrs; **Teacher emotional ability:**1-7; **Self-efficacy factors:** 1-9. **. Correlation is significant at the 0.01 level (2-tailed).*. Correlation is significant at the 0.05 level (2-tailed).

Table 5.4.11. Geography teacher characteristics: Bivariate correlations between maths teacher characteristics and measures showing a significant effect of **geography classroom at time 1** (N)

Construct	1	2	3	4	5	6	7	8
1. Geography primary school achievement	1 (220)							
2. Geography classroom environment	**.171*** **(193)**	1 (214)						
3. Geography classroom student-teacher relations	**.188**** **(193)**	**.912**** **(214)**	1 (214)					
4. Geography classroom chaos	**.198**** **(197)**	**.190**** **(214)**	**.187**** **(214)**	1 (218)				
5. Years of teaching experience	.032 (175)	-.017 (171)	.008 (171)	**-.153*** **(174)**	1 (209)			
6. Teacher emotional ability	**-.271**** **(175)**	**-.259**** **(171)**	**-.318**** **(171)**	-.096 (174)	**-.606**** **(209)**	1 (209)		
7. Teacher self-efficacy in student engagement	**.221**** **(175)**	**.172*** **(171)**	**.198**** **(171)**	**.298**** **(174)**	**-.589**** **(209)**	**-.222**** **(209)**	1 (209)	
8. Teacher self-efficacy instructional strategies	**.279**** **(175)**	**.169*** **(171)**	**.220**** **(171)**	**.267**** **(174)**	-.059 (209)	**-.652**** **(209)**	**.834**** **(209)**	1 (209)
9. Teacher self-efficacy in classroom management	**.271**** **(175)**	**.196*** **(171)**	**.233**** **(171)**	**.320**** **(174)**	**-.433**** **(209)**	**-.359**** **(209)**	**.946**** **(209)**	**.886**** **(209)**

Scale: Geography primary school achievement 1-5; **Classroom environment:** 0-3; **Student-teacher relations:** 0-3; **Classroom chaos:** 0-1; **Teaching experience:** 12-40 yrs; **Teacher emotional ability:**1-7;**Self-efficacy factors:** 1-9. **. Correlation is significant at the 0.01 level (2-tailed).*. Correlation is significant at the 0.05 level (2-tailed).

Table 5.4.12. Geography teacher characteristics: Bivariate correlations between maths teacher characteristics and measures showing a significant effect of **geography classroom at time 2** (N)

Construct	1	2	3	4	5	6	7	9
1. Geography achievement year 5	1 (225)							
2. Geography performance	**.379**** **(198)**	1 (218)						
3. Geography classroom student-teacher relations	**.191**** **(199)**	.126 (213)	1 (219)					
4. Geography environment	**.219**** **(189)**	**.228**** **(204)**	**.352**** **(209)**	1 (210)				
5. Years of teaching experience	-.131 (179)	.080 (181)	**-.164*** **(182)**	**-.189*** **(175)**	1 (209)			
6. Teacher emotional ability	-.068 (179)	**-.286**** **(181)**	-.030 (182)	**-.224**** **(175)**	**-.606**** **(209)**	1 (209)		
7. Teacher self-efficacy in student engagement	**.256**** **(179)**	**.267**** **(181)**	**.258**** **(182)**	**.423**** **(175)**	**-.589**** **(209)**	**-.222**** **(209)**	1 (209)	
8. Teacher self-efficacy in instructional strategies	**.214**** **(179)**	**.381**** **(181)**	**.197**** **(182)**	**.391**** **(175)**	-.059 (209)	**-.652**** **(209)**	**.834**** **(209)**	1 (209)
9. Teacher self-efficacy in classroom management	**.216**** **(179)**	**.291**** **(181)**	**.252**** **(182)**	**.417**** **(175)**	**-.433**** **(209)**	**-.359**** **(209)**	**.946**** **(209)**	**.886**** **(209)**

Scale: Geography performance 0-37; **Student-teacher relations:** 0-3; **Geography environment:** 1-4; **Teaching experience:** 12-40 yrs; **Teacher emotional ability:**1-7;**Self-efficacy factors:** 1-9. **. Correlation is significant at the 0.01 level (2-tailed).*. Correlation is significant at the 0.05 level (2-tailed).

Table 5.4.13. Geography teacher characteristics: Bivariate correlations between maths teacher characteristics and measures showing a significant effect of **geography classroom at time 3** (N)

Construct	1	2	3	4	5	6	7	8
1. Geography performance	1 (224)							
2. Geography classroom environment	.107 (217)	1 (220)						
3. Geography classroom student-teacher relations	**.132*** **(220)**	**.905**** **(219)**	1 (223)					
4. Geography environment	**.162*** **(214)**	**.423**** **(212)**	**.382**** **(216)**	1 (216)				
5. Years of teaching experience	.082 (188)	-.009 (186)	-.070 (189)	**-.267**** **(184)**	1 (209)			
6. Teacher emotional ability	**-.258**** **(188)**	**-.240**** **(186)**	-.136 (189)	-.135 (184)	**-.606**** **(209)**	1 (209)		
7. Teacher self-efficacy student engagement	**.222**** **(188)**	**.250**** **(186)**	**.221**** **(189)**	**.402**** **(184)**	**-.589**** **(209)**	**-.222**** **(209)**	1 (209)	
8. Teacher self-efficacy in instructional strategies	**.305**** **(188)**	**.272**** **(186)**	**.198**** **(189)**	**.298**** **(184)**	-.059 (209)	**-.652**** **(209)**	**.834**** **(209)**	1 (209)
9. Teacher self-efficacy in classroom management	**.226**** **(188)**	**.269**** **(186)**	**.224**** **(189)**	**.390**** **(184)**	**-.433**** **(209)**	**-.359**** **(209)**	**.946**** **(209)**	**.886**** **(209)**

Scale: Geography performance 0-37; **Classroom environment:** 0-3; **Student-teacher relations:** 0-3; **Geography environment:** 1-4; **Geography anxiety:** 1-5; **Teaching experience:** 12-40; **Teacher emotional ability:**1-7;**Self-efficacy factors:** 1-9. **. Correlation is significant at the 0.01 level (2-tailed).*. Correlation is significant at the 0.05 level (2-tailed

Appendix 6

Supplementary materials for Chapter 7

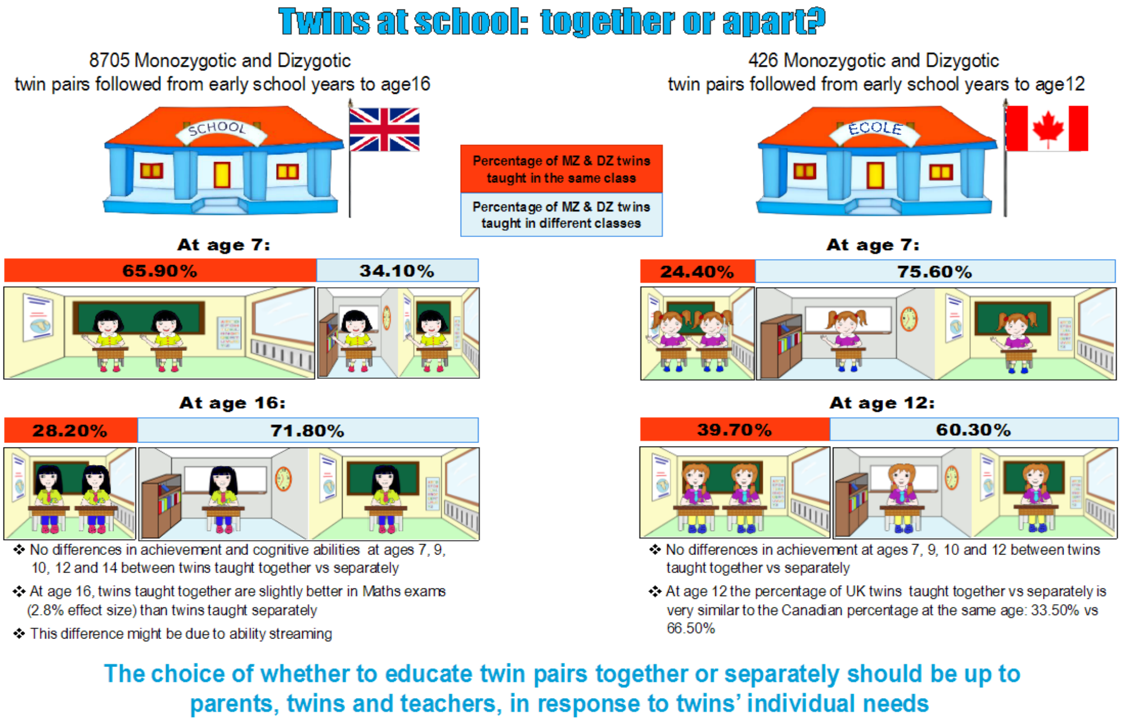

Figure 6.1. Graphical Abstract

Methods

Sample Description

Twins Early Development Study (TEDS). The Twins Early Development Study (TEDS; Howarth, Davis & Plomin, 2013) is an on-going longitudinal study of a representative sample of twins born in England and Wales between 1994 and 1996. The sample consists of three cohorts of families who were initially recruited via the Office of National Statistics (ONS) who contacted the families of all live twin births in England and Wales between January 1994 and December 1996. The first data collection happened when the twins were 18 months old, when demographic data were collected. Since then, families were invited to take part in various studies periodically at ages 2, 3, 4, 7, 8, 9, 10, 12, 14, 16 and 18 years and continuing. Zygosity was established using a parent-reported questionnaire of physical similarity, which is over 95% accurate when compared to DNA testing (Price et al., 2000). For cases where zygosity was unclear, DNA testing was conducted. In taking part, participants were rewarded with gift vouchers and given the opportunity to be entered into monthly prize draws. All participants continue to have access to a 24-hour phone line if they have any questions regarding the study. They also receive leaflets annually, which provide updates on recent research using their data. The total sample consists of 19,522 individuals (3395 monozygotic and 6366 dizygotic twin pairs). To ensure fair comparisons on test scores, 2,112 participants were excluded from analyses on the basis of medical issues and if English was spoken as a second language. For this study following exclusions, just one twin from each pair was selected at random from the remaining sample of 17,410 individual twins (*N*=8,705 pairs).

Quebec Newborn Twin Study (QNTS). The Quebec Newborn Twin

Study (QNTS) is an ongoing prospective longitudinal investigation of a birth cohort of twins that started in the Province of Quebec, Canada, between 1 April 1995 and 31 December 1998. All parents living in the Greater Montreal Area were asked to enroll with their twins in the QNTS. Of 989 families contacted, 672 agreed to participate (68%). Parents were contacted by letter and by phone; laboratory appointments were scheduled for when the twins were five months old (corrected for gestational duration). During the 4–5-hour morning laboratory visit, the mother and her twins were assessed on a number of psychophysiological, cognitive and Behavioural measures. Two weeks later, the families were also visited at home, where the mother was interviewed and both parents filled out questionnaires. These families were seen in the laboratory and in their home between June 1996 and November 1998. The assessments were done in French or English according to the language of the respondent. A broad range of social, demographic, health, and Behavioural data were obtained. Zygosity was ascertained by assessment of physical similarity of twins through aggregation of independent tester ratings using the short version of the Zygosity Questionnaire for Young Twins (Goldsmith, 1991). In addition, DNA was extracted through mouth swabs collected by mothers for 31.3% of the pairs selected at random. DNA-based zygosity was determined using 8–10 polymorphic micro-satellite markers. A comparison of the two methods indicated a concordance of 92%. Taking into account the chorionicity data, available from the twins' medical files, in addition to physical similarity led to an increased concordance rate of 96% (Forget-Dubois et al., 2003).

Measures

The TEDS measures and time of data collection are summarized below in Tables 6.1.1, 6.1.2 and 6.1.3.

Table 6.1.1. Measures description and N for achievement and verbal ability for the UK sample

Age	UK sample n		Description
Achievement			
	Maths	**English**	Teacher reported National Curriculum levels for each subject based on the published versions at the time of each study. Levels range from 1-4, 1-5, and 1-7 depending on guidelines at that time. For current versions see https://www.gov.uk/government/collections/national-curriculum
7 years	5454	5571	
9 years	2594	2602	
10 years	2719	2730	
12 years	3595	3623	
14 years	444	461	
16 years	1634	1635	General certificate of secondary education (GCSE) qualifications. Internationally recognised externally assessed exams taken for specific subjects at age 16. The exams are graded A* to G with A* being the highest. Used here were maths, English, English language and English literature. For assessment guidelines https://www.gov.uk/government/consultations/gcse-subject-content-and-assessment-objectives
Verbal Ability			
7 years	4434		WISC subtests were used to make composite measures of verbal ability: similarities and vocabulary were used at age 7; word quiz; and general knowledge tests were used at ages 9, 10 and 12 accordingly. At age 12 the branching rule was changed so participants enter the test at a higher level. At age 14 the vocabulary test used in the 12 Year study was revised: the first three items were removed to shorten the test; remaining items were reordered to improve difficulty; the branching was removed; and the discontinue rule modified.
9 years	2981		
10 years	2267		
12 years	4200		
14 years	3091		

Table 6.1.2. Measures description and N for non-verbal ability for the U

Age	UK sample n	Description
		Non-verbal Ability
7 years	4462	A non-verbal composite comprised of WISC picture completion subscale and McCarthy conceptual groups test.
9 years	2910	Cognitive Abilities Test 3 figure classification and analogies were used to make a non-verbal composite.
10 years	2245	A composite non-verbal measure comprised of WISC III picture completion subtest and Raven's progressive matrices was used for both age 10 and 12. The Raven's task was revised for age 12 to shorten the test and increase difficulty.
12 years	4052	
14 years	2635	Raven's standard progressive matrices was used at age 14 the test was expanded to include the even numbered items which were removed at age 12.
		General Cognitive Ability (G)
7 years	4428	General cognitive ability composite derived at each age from the verbal and non-verbal tests.
9 years	2906	
10 years	2230	
12 years	4066	
14 years	2628	
		Reading ability
7 years	4408	TOWRE tests of sight word efficiency and phonemic decoding efficiency (word and non-word tests) were used for ages 7 and 12. Peabody Individual Achievement test (PIAT) of reading comprehension used for age 10.
10 years	2530	
12 years	4069	

Table 6.1.3. Measures description and N motivational constructs (Cronbach's alpha) for the UK

Age	UK sample n			Description
	Subject	**SPA n (Alpha)**	**Enjoy n (Alpha)**	Self-perceptions of ability (SPA) and enjoyment of specific subjects were obtained by asking participants 'how much do you like…' and 'how good do you think you are at…' for 3 aspects of the subject. Participants respond using a 5 point scale where 1 = very good or like very much and 5 = not good at all, and don't like at all. Composite scores for overall academic motivation at ages 9 and 12 were derived from SPA and enjoyment for all four subjects at each age.
9 years	Maths	3050 (.814)	2967 (.856)	
	English	3081 (.611)	3026 (.659)	
	Science	3066 (.651)	3014 (.708)	
	PE	3059 (.706)	3005 (.695)	
12 years	Maths	5365 (.859)	5347 (.870)	
	English	5360 (.695)	5353 (.698)	
	Science	5349 (.707)	5355 (.729)	
	PE	5372 (.801)	5372 (.779)	

Cronbach's Alpha were conducted in the present sample on one twin selected randomly from each pair, following exclusions

Table 6.2. School achievement and cognitive ability: Means, standard deviations (SD) and N for Quebec twin pairs taught by the same or different teachers at age 7, for the whole sample; by sex and zygosity; and by zygosity

Construct	Teacher	Whole sample	MZm	DZm	MZf	DZf	DZos	MZ	DZ
Reading	Different	3.27 (1.21) n=162	-0.18 (1.01) n=49	-0.17 (0.98) n=42	0.26 (0.83) n=43	-0.27 (1.04) n=33	-0.03 (1.01) n=73	0.03 (0.95) n=92	-0.12 (1.00) n=148
	Same	3.39 (1.31) n=59	-0.37 (1.10) n=15	0.23 (0.98) n=12	0.02 (0.89) n=23	-0.07 (1.14) n=10	0.20 (1.07) n=91	-0.13 (0.99) n=38	0.14 (1.04) n=40
Writing	Different	3.12 (1.15) n=161	-0.10 (0.99) n=49	-0.26 (0.97) n=41	0.32 (0.83) n=43	-0.25 (1.04) n=33	-0.10 (1.07) n=73	0.10 (0.94) n=92	-0.18 (1.03) n=147
	Same	3.31 (1.15) n=59	-0.31 (1.03) n=15	0.25 (0.91) n=12	0.28 (0.85) n=23	-0.19 (1.25) n=10	0.26 (0.87) n=18	0.05 (0.96) n=38	0.14 (0.98) n=40
Maths	Different	3.50 (1.09) n=161	0.10 (0.98) n=49	0.02 (1.10) n=41	-0.04 (0.96) n=43	-0.27 (0.99) n=33	-0.01 (1.02) n=73	0.03 (0.97) n=92	-0.06 (1.04) n=147
	Same	1.69 (1.09) n=59	-0.18 (1.12) n=15	0.06 (0.96) n=12	0.21 (0.92) n=23	-0.03 (0.91) n=10	0.35 (1.10) n=18	0.06 (1.00) n=38	0.17 (1.00) n=40
General	Different	3.30 (1.11) n=162	-0.07 (0.96) n=49	-0.12 (1.01) n=42	0.12 (0.84) n=43	-0.33 (1.15) n=33	0.04 (1.07) n=73	0.02 (0.91) n=92	-0.09 (1.07) n=148
	Same	3.53 (1.12) n=59	-0.42 (1.12) n=15	0.18 (0.90) n=12	0.18 (0.86) n=23	-0.08 (1.20) n=10	0.46 (1.03) n=18	-0.06 (1.00) n=38	0.24 (1.04) n=40

Means for whole sample are raw and include outliers; means for zygosity and sex and zygosity groups are regressed on age with outliers removed. MZm = monozygotic male; DZm = dizygotic male; MZf = monozygotic female; DZf = dizygotic female; DZos = dizygotic opposite sex; MZ = all monozygotic; DZ = all dizygotic. Significant results in bold at $p \leq .001$ (.05/76).

Table 6.3. School achievement and cognitive ability: Means, standard deviations (SD) and N for UK twin pairs taught by the same or different teachers at age 7, for the whole sample; by sex and zygosity; and by zygosity

Construct	Teacher	Whole sample	MZm	DZm	MZf	DZf	DZos	MZ	DZ
Maths	Different	-0.02 (1.03) n=1908	-0.02 (1.01) n=324	-0.02 (0.99) n=334	-0.05 (0.90) n=341	0.02 (0.94) n=297	0.09 (0.96) n=590	-0.04 (0.95) n=665	0.04 (0.97) n=1221
	Same	0.03 (0.98) n=3600	0.03 (0.96) n=575	0.12 (0.99) n=553	0.01 (0.85) n=688	-0.02 (0.90) n=635	0.12 (0.92) n=1117	0.02 (0.90) n=1263	0.08 (0.93) n=2305
English	Different	-0.03 (1.04) n=1921	-0.17 (0.95) n=322	-0.17 (0.95) n=329	0.10 (0.92) n=345	0.15 (0.95) n=303	0.13 (0.95) n=590	-0.03 (0.94) n=667	0.05 (0.96) n=1222
	Same	0.04 (0.96) n=3615	-0.11 (0.92) n=574	0.03 (0.93) n=555	0.12 (0.87) n=690	0.16 (0.87) n=636	0.10 (0.91) n=1127	0.02 (0.90) n=1264	0.10 (0.90) n=2318
Non-verbal	Different	0.00 (0.99) n=1568	-0.10 (0.99) n=280	-0.05 (1.02) n=268	-0.05 (0.97) n=291	0.10 (0.98) n=257	0.08 (0.96) n=468	-0.07 (0.98) n=571	0.05 (0.98) n=993
	Same	0.03 (1.00) n=2907	-0.02 (-0.02) n=464	-0.01 (-0.01) n=438	0.03 (0.03) n=579	0.13 (0.13) n=500	0.05 (0.05) n=917	0.01 (1.00) n=1043	0.06 (0.97) n=1855
Verbal	Different	0.05 (1.00) n=1558	-0.06 (1.00) n=276	-0.01 (0.91) n=265	-0.06 (0.99) n=291	0.17 (0.91) n=254	0.13 (1.02) n=464	-0.06 (0.99) n=567	0.10 (0.97) n=983
	Same	0.00 (1.00) n=2896	-0.03 (0.99) n=458	-0.05 (1.01) n=439	-0.03 (0.92) N=576	-0.03 (0.97) n=501	0.06 (0.98) n=910	-0.03 (0.95) n=1034	0.01 (0.99) n=1850
g	Different	0.03 (1.00) n=1558	-0.10 (0.98) n=278	-0.05 (0.98) n=265	-0.07 (0.97) n=291	0.17 (0.96) n=254	0.12 (1.01) n=464	-0.08 (0.98) n=569	0.09 (0.99) n=983
	Same	0.02 (0.99) n=2890	-0.04 (0.98) n=457	-0.04 (1.00) n=437	0.00 (0.93) n=574	0.05 (0.97) n=499	0.08 (0.95) n=909	-0.02 (0.95) n=1031	0.04 (0.97) n=1845
Reading	Different	0.00 (1.01) n1546	-0.17 (1.08) n=276	-0.05 (1.02) n=266	0.00 (0.95) n=285	0.07 (0.95) n=255	0.10 (1.02) n=463	-0.08 (1.02) n=561	0.05 (1.00) n=984
	Same	0.03 (0.99) n=2864	-0.06 (1.02) n=457	0.03 (1.01) n=439	0.07 (0.96) n=573	-0.01 (0.96) n=490	0.08 (0.99) n=904	0.01 (0.99) n=1030	0.04(0.99) n=1833

Means for whole sample are raw and include outliers; means for zygosity and sex and zygosity groups are regressed on age with outliers removed. MZm = monozygotic male; DZm = dizygotic male; MZf = monozygotic female; DZf = dizygotic female; DZos = dizygotic opposite sex; MZ = all monozygotic; DZ = all dizygotic. Significant results in bold at $p \leq .001$ (.05/76).

Table 6.4. School achievement and cognitive ability: Means, standard deviations (SD) and N for Quebec twin pairs taught by the same or different teachers at age 9, for the whole sample; by sex and zygosity; and by zygosity

Construct	Teacher	Whole sample	MZm	DZm	MZf	DZf	DZos	MZ	DZ
Reading	Different	3.19 (1.13) n=223	-0.17 (1.09) n=57	0.00 (1.14) n=43	0.16 (0.97) n=65	-0.11 (0.92) n=40	0.09 (0.93) n=84	0.01 (1.04) n=122	0.02 (0.98) n=167
	Same	3.04 (1.07) n=70	-0.25 (1.11) n=16	0.32 (1.08) n=14	0.02 (1.00) n=21	0.16 (1.12) n=13	-0.19 (0.79) n=24	-0.10 (1.05) n=37	0.04 (0.97) n=51
Writing	Different	3.06 (1.22) n=222	-0.18 (1.05) n=57	-0.10 (1.11) n=42	0.31 (0.96) n=65	-0.02 (0.97) n=40	0.10 (1.03) n=84	0.08 (1.03) n=122	0.02 (1.03) n=166
	Same	2.74 (1.11) n=70	-0.37 (1.05) n=16	0.16 (1.00) n=14	0.01 (1.06) n=21	-0.02 (0.99) n=13	-0.13 (0.84) n=24	-0.15 (1.06) n=37	-0.02 (0.91) n=51
Maths	Different	3.18 (1.17) n=221	0.09 (0.99) n=57	0.86 (1.09) n=42	0.03 (0.93) n=65	-0.43 (1.05) n=40	0.04 (0.96) n=83	0.06 (0.96) n=122	-0.06 (1.03) n=165
	Same	3.25 (1.05) n=69	-0.03 (0.99) n=16	0.51 (1.23) n=14	0.04 (0.89) n=21	-0.12 (0.73) n=11	0.12 (0.85) n=24	0.01 (0.92) n=37	0.18 (0.96) n=49
Science	Different	3.17 (1.07) n=222	-0.06 (1.19) n=56	0.13 (1.21) n=41	0.13 (1.03) n=63	-0.10 (0.93) n=40	0.03 (0.93) n=77	0.04 (0.11) n=119	0.02 (1.01) n=158
	Same	3.14 (0.87) n=70	-0.17 (0.90) n=16	0.35 (1.20) n=14	-0.01 (1.09) n=21	-0.15 (0.76) n=12	-0.04 (0.74) n=24	-0.08 (1.00) n=37	0.04 (0.90) n=50
General	Different	3.31 (0.98) n=213	-0.13 (1.11) n=57	0.05 (1.13) n=43	0.14 (0.99) n=65	-14.00 (0.95) N=39	-0.03 (0.97) n=84	0.01 (1.05) n=122	-0.03 (1.01) n=166
	Same	3.19 (0.83) n=70	-0.31 (0.97) n=16	0.27 (1.07) n=14	0.14 (0.88) n=21	0.07 (0.98) n=13	0.04 (0.66) n=24	-0.06 (0.93) n=37	0.11 (0.86) n=51

Means for whole sample are raw and include outliers; means for zygosity and sex and zygosity groups are regressed on age with outliers removed. MZm = monozygotic male; DZm = dizygotic male; MZf = monozygotic female; DZf = dizygotic female; DZos = dizygotic opposite sex; MZ = all monozygotic; DZ = all dizygotic. Significant results in bold at $p \leq .001$ (.05/76).

Table 6.5. School achievement and cognitive ability: Means, standard deviations (SD) and N for UK twin pairs taught by the same or different teachers at age 9, for the whole sample; by sex and zygosity; and by zygosity

Construct	Teacher	Whole sample	MZm	DZm	MZf	DZf	DZos	MZ	DZ
Maths	Different	-0.03 (1.00) n=1079	-0.03 (1.00) n=170	0.08 (1.03) n=169	-0.15 (0.98) n=206	-0.05 (0.94) n=188	0.02 (0.99) n=343	-0.09 (0.99) n=376	0.01 (0.99) n=700
	Same	0.03 (0.99) n=1529	-0.01 (1.03) n=252	0.09 (0.93) n=224	-0.06 (0.95) n=313	0.05 (0.94) n=263	0.06 (0.95) n=466	-0.03 (0.99) n=565	0.06 (0.94) n=953
English	Different	-0.05 (1.03) n=1083	-0.22 (1.02) n=167	-0.17 (0.99) n=170	0.11 (0.94) n=203	0.09 (1.00) n=187	0.01 (0.98) n=345	-0.04 (0.99) n=370	-0.01 (0.99) n=702
	Same	0.04 (0.97) n=1545	-0.14 (0.92) n=248	-0.02 (0.93) n=230	0.09 (0.93) n=317	0.20 (0.94) n=264	0.06 (0.92) n=471	-0.01 (0.93) n=565	0.08 (0.93) n=965
Non-verbal	Different	-0.02 (1.00) n=1198	-0.12 (1.04) n=193	-0.03 (1.01) n=181	-0.02 (0.99) n=242	0.04 (0.98) n=199	0.05 (0.95) n=380	-0.06 (1.01) n=435	0.02 (0.97) n=760
	Same	0.03 (0.99) n=1717	0.01 (0.98) n=266	0.19 (0.98) n=255	0.04 (0.97) n=374	-0.10 (1.07) n=304	0.03 (0.94) n=516	0.03 (0.98) n=640	0.03 (0.99) n=1075
Verbal	Different	-0.04 (0.97) n=1216	-0.08 (0.95) n=198	-0.02 (0.93) n=182	-0.05 (0.92) n=247	-0.09 (1.03) n=203	0.03 (0.97) n=384	-0.06 (0.94) n=445	-0.01 (0.98) n=769
	Same	0.05 (0.96) n=1769	0.02 (0.99) n=278	0.17 (0.97) n=264	-0.05 (0.97) n=382	0.03 (0.95) n=313	0.09 (0.93) n=530	-0.02 (0.98) n=660	0.10 (0.94) n=1107
g	Different	0.03 (1.00) n=1558	-0.11 (0.99) n=192	-0.04 (0.95) n=179	-0.05 (0.96) n=243	-0.03 (1.04) n=200	0.04 (0.96) n=378	-0.08 (0.97) n=435	0.00 (0.98) n=757
	Same	0.02 (1.00) n=2890	0.02 (0.99) n=264	0.23 (0.95) n=255	-0.01 (0.96) n=374	-0.05 (1.03) n=305	0.07 (0.91) n=516	0.00 (0.97) n=638	0.07 (0.96) n=1076

Means for whole sample are raw and include outliers; means for zygosity and sex and zygosity groups are regressed on age with outliers removed. MZm = monozygotic male; DZm = dizygotic male; MZf = monozygotic female; DZf = dizygotic female; DZos = dizygotic opposite sex; MZ = all monozygotic; DZ = all dizygotic. Significant results in bold at $p \leq .001$ (.05/76).

Table 6.6. School achievement and cognitive ability: Means, standard deviations (SD) and N for Quebec twin pairs taught by the same or different teachers at age 10, for the whole sample; by sex and zygosity; and by zygosity

Construct	Teacher	Whole sample	MZm	DZm	MZf	DZf	DZos	MZ	DZ
Reading	Different	3.07 (1.12) n=213	-0.22 (0.96) n=36	-0.04 (1.05) n=35	-0.10 (0.87) n=44	-0.13 (1.02) n=37	0.23 (0.99) n=59	-0.15 (0.91) n=80	0.05 (1.02) n=131
	Same	3.03 (1.21) n=88	-0.33 (1.06) n=16	0.16 (1.19) n=12	0.37 (0.97) n=23	-0.63 (0.80) n=11	-0.05 (1.03) n=25	0.08 (1.06) n=39	-0.13 (1.04) n=48
Writing	Different	2.87 (1.25) n=215	-0.31 (1.07) n=58	-0.19 (1.04) n=43	0.12 (0.84) n=53	-0.16 (0.87) n=43	0.20 (0.96) n=76	-0.11 (0.99) n=111	0.00 (0.97) n=162
	Same	2.90 (1.31) n=88	-0.46 (0.89) n=23	0.38 (0.91) n=16	0.46 (0.93) n=31	-0.19 (0.93) n=17	0.06 (1.14) n=28	0.07 (1.01) n=54	0.08 (1.03) n=61
Maths	Different	3.12 (1.17) n=211	-0.12 (1.05) n=36	-0.07 (1.09) n=34	0.03 (0.91) n=44	-0.26 (0.97) n=37	0.15 (0.93) n=58	-0.04 (0.97) n=80	-0.03 (0.99) n=129
	Same	3.15 (1.33) n=87	-0.25 (1.06) n=16	0.48 (1.17) n=12	0.17 (1.11) n=23	-0.92 (0.81) n=11	0.23 (1.05) n=24	0.00 (1.09) n=39	0.03 (1.14) n=47
General	Different	3.11 (1.17) n=208	-0.33 (1.05) n=57	-0.02 (1.10) n=41	0.05 (0.86) n=51	-0.12 (1.01) n=42	0.23 (0.88) n=75	-0.15 (0.98) n=108	0.07 (0.98) n=158
	Same	3.07 (1.27) n=88	-0.38 (0.91) n=23	0.15 (1.14) n=16	0.34 (0.95) n=31	-0.44 (1.06) n=17	0.24 (1.05) n=28	0.04 (0.99) n=54	0.03 (1.10) n=61

Means for whole sample are raw and include outliers; means for zygosity and sex and zygosity groups are regressed on age with outliers removed. MZm = monozygotic male; DZm = dizygotic male; MZf = monozygotic female; DZf = dizygotic female; DZos = dizygotic opposite sex; MZ = all monozygotic; DZ = all dizygotic. Significant results in bold at $p \leq .001$ (.05/76).

Table 6.7. School achievement and cognitive ability: Means, standard deviations (SD) and N for UK twin pairs taught by the same or different teachers at age 10, for the whole sample; by sex and zygosity; and by zygosity

Construct	Teacher	Whole sample	MZm	DZm	MZf	DZf	DZos	MZ	DZ
Maths	Different	-0.03 (1.03) n=1256	0.06 (0.98) n=191	0.07 (0.94) n=200	-0.12 (0.89) n=239	0.00 (1.01) n=215	0.04 (0.99) n=395	-0.04 (0.93) n=430	0.04 (0.98) n=810
	Same	0.04 (0.98) n=1489	0.06 (0.99) n=233	0.21 (1.01) n=214	0.03 (0.92) n=309	-0.01 (0.94) n=245	0.06 (0.92) n=478	0.05 (0.95) n=542	0.08 (0.95) n=937
English	Different	0.00 (1.05) n=1258	-0.10 (1.01) n=188	-0.08 (0.97) n=204	0.08 (0.95) n=238	0.15 (1.07) n=221	0.05 (0.99) n=394	0.00 (0.98) n=426	0.04 (1.01) n=819
	Same	0.04 (0.96) n=1492	-0.19 (0.99) n=238	0.06 (0.97) n=216	0.17 (0.91) n=310	0.09 (0.87) n=245	0.08 (0.93) n=476	0.01 (0.96) n=548	0.08 (0.92) n=937
Non-verbal	Different	0.01 (1.00) n=1030	0.06 (0.99) n=148	0.12 (1.00) n=150	-0.05 (0.96) n=207	-0.03 (1.04) n=193	0.03 (0.94) n=328	-0.01 (0.97) n=355	0.04 (0.98) n=671
	Same	0.05 (0.95) n=1223	0.07 (0.98) n=178	0.22 (0.91) n=169	0.02 (0.95) n=280	0.05 (0.97) n=205	0.02 (0.90) n=387	0.04 (0.96) n=458	0.07 (0.92) n=761
Verbal	Different	0.01 (1.02) n=1040	0.11 (1.07) n=148	0.15 (1.10) n=152	-0.11 (0.97) n=210	-0.07 (1.01) n=196	0.04 (0.95) n=333	-0.02 (1.01) n=358	0.03 (1.01) n=681
	Same	0.02 (0.95) n=1228	0.11 (0.95) n=179	0.34 (0.87) n=170	-0.13 (0.96) n=281	-0.12 (1.00) n=208	0.04 (0.93) n=390	-0.04 (0.96) n=460	0.06 (0.95) n=768
g	Different	0.01 (1.01) n=1024	0.08 (1.07) n=146	0.16 (1.08) n=149	-0.11 (0.96) n=208	-0.07 (1.04) n=193	0.04 (0.95) n=327	-0.03 (1.01) n=354	0.04 (1.01) n=669
	Same	0.04 (0.96) n=1209	0.10 (0.97) n=176	0.31 (0.85) n=168	-0.07 (0.97) n=277	-0.04 (1.00) n=204	0.03 (0.92) n=382	-0.01 (0.97) n=453	0.08 (0.94) n=754
Reading	Different	47.19 (13.14) n=1156	-0.10 (1.12) n=170	0.06 (1.01) n=177	-0.02 (0.92) n=230	0.10 (0.95) n=209	0.11 (0.96) n=368	-0.05 (1.01) n=400	0.10 (0.97) n=754
	Same	47.17 (12.75) n=1380	0.03 (1.00) n=208	0.23 (0.90) n=189	-0.02 (0.92) n=310	0.04 (0.99) n=231	0.04 (0.94) n=438	0.00 (0.95) n=518	0.08 (0.95) n=858

Means for whole sample are raw and include outliers; means for zygosity and sex and zygosity groups are regressed on age with outliers removed. MZm = monozygotic male; DZm = dizygotic male; MZf = monozygotic female; DZf = dizygotic female; DZos = dizygotic opposite sex; MZ = all monozygotic; DZ = all dizygotic. Significant results in bold at $p \leq .001$ (.05/76).

Table 6.8. School achievement and cognitive ability: Means, standard deviations (SD) and N for Quebec twin pairs taught by the same or different teachers at age 12, for the whole sample; by sex and zygosity; and by zygosity

Construct	Teacher	Whole sample	MZm	DZm	MZf	DZf	DZos	MZ	DZ
Reading	Different	3.17 (1.16) n=177	-0.05 (1.10) n=30	-0.45 (0.93) n=27	0.09 (0.88) n=39	-0.02 (0.82) n=31	-0.07 (1.05) n=59	0.03 (0.98) n=69	-0.13 (0.97) n=116
	Same	3.43 (1.10) n=118	-0.44 (0.69) n=19	0.39 (0.87) n=19	0.30 (0.85) n=31	0.38 (1.00) n=17	0.18 (0.95) n=38	0.02 (0.87) n=50	0.28 (0.94) n=74
Writing	Different	3.00 (1.29) n=178	-0.38 (0.91) n=30	-0.44 (1.15) n=27	0.10 (0.94) n=39	-0.03 (0.87) n=31	0.02 (1.06) n=59	-0.11 (0.95) n=69	-0.09 (1.04) n=116
	Same	3.25 (1.25) n=118	-0.51 (0.77) n=19	0.30 (0.92) n=19	0.43 (0.85) n=31	0.45 (1.02) n=17	0.15 (1.00) n=38	0.07 (0.93) n=50	0.26 (0.98) n=74
Maths	Different	3.22 (1.21) n=175	-0.13 (1.03) n=29	-0.24 (1.04) n=27	0.07 (0.85) n=38	-0.34 (0.93) n=31	-0.07 (1.07) n=58	-0.01 (0.93) n=67	-0.17 (1.02) n=115
	Same	3.59 (1.13) n=114	-0.22 (0.69) n=17	0.49 (0.92) n=19	0.24 (0.89) n=29	0.24 (1.09) n=17	0.24 (0.92) n=37	0.07 (0.84) n=46	0.31 (0.95) n=73
General	Different	3.10 (1.25) n=174	-0.26 (1.07) n=29	-0.44 (1.00) n=27	0.18 (0.86) n=39	-0.18 (1.01) n=30	-0.06 (1.12) n=57	-0.01 (0.97) n=68	-0.17 (1.06) n=113
	Same	3.51 (1.15) n=118	-0.38 (0.75) n=19	0.40 (0.89) n=19	0.39 (0.80) n=31	0.19 (1.12) n=17	0.38 (1.02) n=38	0.10 (0.86) n=50	0.34 (1.00) n=74

Means for whole sample are raw and include outliers; means for zygosity and sex and zygosity groups are regressed on age with outliers removed. MZm = monozygotic male; DZm = dizygotic male; MZf = monozygotic female; DZf = dizygotic female; DZos = dizygotic opposite sex; MZ = all monozygotic; DZ = all dizygotic. Significant results in bold at $p \leq .001$ (.05/76).

Table 6.9. School achievement and cognitive ability: Means, standard deviations (SD) and N for the UK twin pairs taught by the same or different teachers at age 12, for the whole sample; by sex and zygosity; and by zygosity

Construct	Teacher	Whole sample	MZm	DZm	MZf	DZf	DZos	MZ	DZ
Maths	Different	0.07 (1.01) n=2086	-0.14 (1.00) n=332	0.02 (0.91) n=323	-0.04 (0.88) n=376	0.01 (0.92) n=336	0.00 (0.92) n=695	-0.08 (0.94) n=708	0.01 (0.92) n=1354
	Same	-0.05 (0.99) n=1561	0.06 (0.87) n=244	-0.01 (0.87) n=226	-0.06 (0.89) n=337	-0.09 (0.89) n=306	0.02 (0.87) n=420	-0.01 (0.88) n=581	-0.02 (0.88) n=952
English	Different	0.05 (1.04) n=2116	-0.20 (0.96) n=344	-0.13 (1.00) n=327	0.10 (0.85) n=381	0.06 (0.92) n=343	-0.02 (0.92) n=691	-0.05 (0.92) n=725	-0.02 (0.94) n=1361
	Same	-0.01 (0.97) n=1571	-0.06 (0.79) n=242	-0.04 (0.80) n=224	0.04 (0.84) n=341	0.09 (0.77) n=313	0.05 (0.86) n=417	0.00 (0.82) n=583	0.04 (0.82) n=954
Non-verbal	Different	0.04 (1.01) n=2580	-0.04 (1.05) n=389	0.07 (0.99) n=370	-0.04 (0.94) n=524	0.06 (0.97) n=428	0.05 (0.95) n=852	-0.04 (0.98) n=913	0.06 (0.96) n=1650
	Same	-0.06 (0.98) n=1502	0.06 (0.95) n=225	0.20 (1.03) n=206	0.00 (0.89) n=363	0.04 (0.91) n=301	-0.02 (0.96) n=394	0.02 (0.91) n=588	0.05 (0.96) n=901
Verbal	Different	0.03 (1.00) n=2654	0.06 (0.98) n=399	0.09 (1.01) n=386	-0.20 (1.01) n=540	-0.03 (1.01) n=448	0.08 (0.97) n=879	-0.09 (1.01) n=939	0.05 (0.99) n=1713
	Same	-0.01 (0.98) n=1549	0.13 (0.96) n=235	0.31 (0.92) n=216	-0.15 (0.96) n=374	-0.02 (0.91) n=310	0.05 (1.03) n=413	-0.04 (0.97) n=609	0.09 (0.98) n=939
g	Different	0.04 (1.01) n=2580	0.01 (1.01) n=390	0.09 (1.00) n=370	-0.15 (0.99) n=529	0.00 (1.02) n=432	0.08 (0.95) n=851	-0.08 (1.00) n=919	0.06 (0.98) n=1653
	Same	-0.04 (0.97) n=1502	0.12 (0.92) n=225	0.30 (0.97) n=208	-0.10 (0.89) n=365	0.02 (0.91) n=301	0.02 (1.00) n=395	-0.01 (0.91) n=590	0.08 (0.97) n=904
Reading	Different	0.02 (1.01) n=2621	-0.04 (1.03) n=398	0.06 (0.93) n=371	-0.02 (0.97) n=523	0.08 (0.97) n=446	0.02 (0.95) n=861	-0.03 (0.99) n=921	0.05 (0.95) n=1678
	Same	-0.06 (0.99) n=1481	-0.02 (0.98) n=231	0.06 (0.92) n=209	-0.04 (0.93) n=357	0.06 (0.92) n=299	0.02 (0.96) n=374	-0.03 (0.95) n=588	0.04 (0.93) n=882

Means for whole sample are raw and include outliers; means for zygosity and sex and zygosity groups are regressed on age with outliers removed. MZm = monozygotic male; DZm = dizygotic male; MZf = monozygotic female; DZf = dizygotic female; DZos = dizygotic opposite sex; MZ = all monozygotic; DZ = all dizygotic. Significant results in bold at $p \leq .001$ (.05/76).

Table 6.10. School achievement and cognitive ability: Means, standard deviations (SD) and N for the UK twin pairs taught by the same or different teachers at age 14, for the whole sample; by sex and zygosity; and by zygosity

Construct	Teacher	Whole sample	MZm	DZm	MZf	DZf	DZos	MZ	DZ
Maths	Different	6.36 (1.66) n=2504	-0.17 (1.19) n=54	-0.04 (1.10) n=52	-0.12 (1.13) n=81	0.03 (0.85) n=57	-0.02 (1.07) n=106	-0.14 (1.15) n=135	-0.01 (1.02) n=215
	Same	6.05 (2.06) n=193	0.24 (0.87) n=16	0.02 (0.91) n=15	0.04 (0.87) n=33	0.34 (0.47) n=27	0.28 (0.78) n=22	0.11 (0.86) n=49	0.24 (0.70) n=64
English	Different	**5.78 (1.47) n=2508**	-0.10 (1.05) n=54	-0.22 (1.09) n=52	-0.15 (1.14) n=80	0.14 (0.94) n=57	-0.05 (1.08) n=106	-0.13 (1.10) n=134	-0.04 (1.05) n=215
	Same	**5.36 (1.83) n=192**	0.20 (0.92) n=14	0.01 (0.95) n=15	0.10 (0.85) n=33	0.29 (0.39) n=27	0.32 (0.77) n=23	0.13 (0.86) n=47	0.24 (0.69) n=65
English reading	Different	5.70 (1.14) n=159	0.04 (0.78) n=24	0.03 (0.81) n=17	0.03 (1.00) n=34	0.00 (0.72) n=24	0.11 (0.96) n=60	0.04 (0.91) n=58	0.07 (0.88) n=101
	Same	5.65 (1.47) n=52	-0.88 (1.65) n=2	0.29 (1.10) n=7	-0.01 (1.26) n=13	0.00 (1.17) n=13	0.29 (0.57) n=16	-0.13 (1.29) n=15	0.18 (0.91) n=36
Non-verbal	Different	13.98 (3.59) n=2377	0.06 (0.98) n=364	0.01 (1.01) n=326	-0.06 (0.94) n=546	0.02 (0.99) n=435	0.03 (0.96) n=699	-0.01 (0.95) n=910	0.02 (0.98) n=1460
	Same	13.95 (3.41) n=266	-0.32 (0.95) n=39	-0.22 (1.12) n=38	0.11 (0.87) n=82	0.14 (0.69) n=55	0.23 (0.93) n=51	-0.03 (0.92) n=121	0.08 (0.92) n=144
Verbal	Different	41.28 (5.32) n=2840	0.03 (0.93) n=410	0.08 (0.93) n=376	0.03 (0.95) n=646	0.13 (0.93) n=522	0.07 (0.96) n=860	0.03 (0.94) n=1056	0.09 (0.94) n=1758
	Same	40.56 (5.65) n=280	-0.13 (1.02) n=41	-0.06 (1.16) n=39	-0.09 (0.98) n=83	0.14 (0.96) n=58	0.00 (1.04) n=56	-0.10 (0.99) n=124	0.04 (1.04) n=153
g	Different	0.03 (1.00) n=2375	0.04 (0.98) n=364	0.05 (0.99) n=328	-0.02 (0.94) n=544	0.07 (0.99) n=433	0.05 (0.98) n=697	0.00 (0.96) n=908	0.06 (0.98) n=1458
	Same	-0.06 (1.01) n=265	-0.18 (0.88) n=38	-0.22 (1.21) n=38	-0.03 (0.98) n=80	0.18 (0.82) n=55	0.13 (0.97) n=51	-0.08 (0.94) n=118	0.06 (1.00) n=144

Means for whole sample are raw and include outliers; means for zygosity and sex and zygosity groups are regressed on age with outliers removed. MZm = monozygotic male; DZm = dizygotic male; MZf = monozygotic female; DZf = dizygotic female; DZos = dizygotic opposite sex; MZ = all monozygotic; DZ = all dizygotic. Significant results in bold at $p \leq .001$ (.05/76).

Table 6.11. School achievement and cognitive ability: Means, standard deviations (SD) and N for the UK twin pairs taught by the same or different teachers at age 16, for the whole sample; by sex and zygosity; and by zygosity

Construct	Teacher	Whole sample	MZm	DZm	MZf	DZf	DZos	MZ	DZ
Maths GCSE	Different	**8.90 (1.41) n=1168**	0.01 (0.88) n=161	**0.01 (0.92) n=179**	-0.05 (0.82) n=169	-0.04 (0.97) n=225	**0.05 (0.95) n=428**	**-0.02 (0.85) n=330**	**0.02 (0.95) n=832**
	Same	**9.40 (1.43) n=474**	0.21 (1.00) n=92	**0.51 (0.80) n=65**	0.25 (0.97) n=136	0.36 (0.84) n=73	**0.47 (0.94 n=106**	**0.24 (0.98) n=228**	**0.45 (0.87) n=244**
English GCSE	Different	**9.06 (1.31) n=1234**	-0.27 (0.92) n=171	-0.18 (0.92) n=193	0.21 (0.81) n=196	0.17 (0.87) n=210	0.07 (0.96) n=451	-0.02 (0.90) n=367	**0.04 (0.93) n=854**
	Same	**9.32 (1.17) n=416**	-0.12 (0.90) n=83	0.27 (0.84) n=51	0.28 (0.82) n=113	0.23 (0.90) n=84	0.32 (0.81) n=83	0.11 (0.87) n=196	**0.27 (0.85) n=218**
English Language GCSE	Different	**8.86 (1.24) n=1226**	-0.26 (0.95) n=170	-0.19 (0.94) n=192	0.18 (0.83) n=195	0.09 (0.90) n=209	0.05 (1.00) n=451	-0.02 (0.92) n=365	**0.01 (0.97) n=852**
	Same	**9.09 (1.20) n=414**	-0.10 (0.93) n=82	0.25 (0.95) n=51	0.22 (0.89) n=112	0.23 (0.93) n=83	0.31 (0.88) n=83	0.08 (0.92) n=194	**0.27 (0.91 n=217**
English Literature GCSE	Different	9.08 (1.21) n=1056	-0.33 (0.98) n=141	-0.07 (0.96) n=156	0.14 (0.89) n=179	0.19 (0.87) n=189	0.12 (0.93) n=382	-0.07 (0.96) n=320	0.10 (0.93) n=727
	Same	9.18 (1.15) n=383	-0.20 (0.99) n=76	0.19 (0.89) n=45	0.15 (0.85) n=105	0.16 (0.99) n=79	0.26 (0.90) n=77	0.01 (0.92) n=181	0.20 (0.93) n=201

Means for whole sample are raw and include outliers; means for zygosity and sex and zygosity groups are regressed on age with outliers removed. MZm = monozygotic male; DZm = dizygotic male; MZf = monozygotic female; DZf = dizygotic female; DZos = dizygotic opposite sex; MZ = all monozygotic; DZ = all dizygotic. Significant results in bold at $p \leq .001$ (.05/76).

Table 6.12. Enjoyment: Means, standard deviations (SD) and N for the UK twin pairs taught by the same or different teachers at age 9, for the whole sample; by sex and zygosity; and by zygosity

Construct	Teacher	Whole sample	MZm	DZm	MZf	DZf	DZos	MZ	DZ
English	Different	3.77 (0.91) n=1210	-0.09 (0.99) n=191	-0.06 (0.95) n=176	0.19 (0.88) n=245	0.09 (0.93) n=198	-0.03 (0.94) n=375	0.07 (0.94) n=436	-0.01 (0.94) n=749
	Same	3.85 (0.86) n=1760	-0.11 (0.93) n=271	-0.02 (0.93) n=259	0.22 (0.91) n=380	0.11 (0.88) n=308	0.10 (0.96) n=523	0.08 (0.93) n=651	0.07 (0.93) n=1090
Maths	Different	3.49 (1.19) n=1209	0.13 (0.99) n=198	0.19 (1.00) n=181	-0.17 (1.03) n=246	-0.22 (1.06) n=201	0.02 (1.01) n=383	-0.04 (1.02) n=444	0.00 (1.03) n=765
	Same	3.53 (1.16) n=1759	0.05 (1.07) n=276	0.19 (1.00) n=262	-0.17 (1.00) n=381	-0.08 (0.94) n=309	0.11 (0.99) n=531	-0.08 (1.03) n=657	0.08 (0.98) n=1102

Table 6.13. Enjoyment: Means, standard deviations (SD) and N for Quebec twin pairs taught by the same or different teachers at age 10, for the whole sample; by sex and zygosity; and by zygosity

Construct	Teacher	Whole sample	MZm	DZm	MZf	DZf	DZos	MZ	DZ
Reading	Different	3.54 (1.29) n=209	0.10 (1.01) n=56	-0.31 (1.10) n=37	-0.04 (1.13) n=52	-0.14 (0.98) n=43	0.12 (0.94) n=72	0.03 (1.06) n=108	-0.06 (1.00) n=152
	Same	3.61 (1.23) n=86	-0.26 (1.08) n=21	-0.11 (1.14) n=16	0.20 (0.97) n=31	0.26 (0.81) n=16	0.10 (1.05) n=25	0.01 (1.03) n=52	0.09 (1.01) n=57
Math	Different	3.38 (1.31) n=209	0.16 (0.85) n=56	-0.05 (0.93) n=37	0.02 (1.10) n=52	-0.02 (1.05) n=43	0.19 (0.95) n=72	0.09 (0.98) n=108	0.07 (0.97) n=152
	Same	3.15 (1.40) n=86	0.01 (1.04) n=21	-0.07 (0.94) n=16	-0.30 (1.20) n=31	-0.09 (0.77) n=16	0.10 (0.98) n=25	-0.17 (1.14) n=52	0.00 (0.90) n=57

Table 6.12 & 6.13: Means for whole sample are raw and include outliers; means for zygosity and sex and zygosity groups are regressed on age with outliers removed. MZm = monozygotic male; DZm = dizygotic male; MZf = monozygotic female; DZf = dizygotic female; DZos = dizygotic opposite sex; MZ = all monozygotic; DZ = all dizygotic. Significant results in bold at $p \leq .001$ (.05/76).

Table 6.14. Enjoyment: Means, standard deviations (SD) and N for Quebec twin pairs taught by the same or different teachers at age 12, for the whole sample; by sex and zygosity; and by zygosity

Construct	Teacher	Whole sample	MZm	DZm	MZf	DZf	DZos	MZ	DZ
Reading	Different	3.11 (1.39) n=215	0.05 (1.08) n=36	-0.43 (1.13) n=33	-0.21 (0.88) n=48	0.17 (0.94) n=36	0.14 (1.02) n=65	-0.10 (0.97) n=84	0.01 (1.05) n=134
	Same	3.36 (1.26) n=143	-0.10 (0.96) n=27	0.10 (1.10) n=24	0.29 (0.90) n=35	0.43 (0.89) n=19	0.01 (0.82) n=38	0.12 (0.94) n=62	0.13 (0.93) n=81
Math	Different	2.81 (1.25) n=215	0.14 (0.98) n=36	-0.41 (1.11) n=33	-0.13 (1.05) n=48	-0.24 (0.94) n=36	-0.07 (0.99) n=65	-0.02 (1.02) n=84	-0.20 (1.01) n=134
	Same	3.19 (1.30) n=143	0.35 (1.04) n=27	0.07 (1.09) n=24	0.09 (1.18) n=35	0.42 (1.03) n=19	0.06 (0.93) n=37	0.20 (1.12) n=62	0.15 (1.00) n=80

Table 6.15. Enjoyment: Means, standard deviations (SD) and N for the UK twin pairs taught by the same or different teachers at age 12, for the whole sample; by sex and zygosity; and by zygosity

Construct	Teacher	Whole sample	MZm	DZm	MZf	DZf	DZos	MZ	DZ
English	Different	**3.43 (0.86) n=3470**	-0.23 (0.97) n=466	-0.29 (1.00) n=449	0.15 (0.99) n=584	0.16 (0.94) n=492	0.02 (0.99) n=1012	-0.02 (1.00) n=1050	-0.01 (0.99) n=1953
	Same	**3.55 (0.83) n=1923**	-0.19 (0.99) n=267	-0.11 (0.93) n=247	0.26 (0.91) n=397	0.32 (0.90) n=339	-0.03 (1.00) n=468	0.07 (0.97) n=664	0.06 (0.97) n=1054
Maths	Different	**3.23 (1.04) n=3473**	0.04 (1.04) n=469	-0.02 (1.02) n=449	-0.08 (0.93) n=585	-0.10 (0.96) n=494	-0.01 (1.02) n=1016	-0.03 (0.98) n=1054	-0.03 (1.01) n=1959
	Same	**3.37 (1.03) n=1922**	0.08 (1.04) n=266	0.09 (0.98) n=247	-0.02 (0.94) n=398	0.05 (0.98) n =340	0.08 (1.03) n=468	0.02 (0.98) n=664	0.07 (1.00) n=1055
Academic	Different	**3.46 (0.68) n=3473**	-0.04 (0.96) n=465	-0.11 (0.96) n=443	-0.01 (1.00) n=582	0.00 (0.95) n=493	0.04 (0.96) n=1010	-0.02 (0.98) n=1047	0.00 (0.96) n=1946
	Same	**3.57 (0.66) n=1923**	-0.01 (0.95) n=264	0.06 (0.97) n=247	0.13 (0.95) n=396	0.19 (0.95) n=339	0.02 (0.98) n=468	0.07 (0.95) n=660	0.08 (0.97) n=1054

Table 6.14 & 6.15: Means for whole sample are raw and include outliers; means for zygosity and sex and zygosity groups are regressed on age with outliers removed. MZm = monozygotic male; DZm = dizygotic male; MZf = monozygotic female; DZf = dizygotic female; DZos = dizygotic opposite sex; MZ = all monozygotic; DZ = all dizygotic. Significant results in bold at $p \leq .001$ (.05/76).

Table 6.16. Self-perceived ability: Means, standard deviations (SD) and N for the UK twin pairs taught by the same or different teachers at age 9, for the whole sample; by sex and zygosity; and by zygosity

Construct	Teacher	Whole sample	MZm	DZm	MZf	DZf	DZos	MZ	DZ
English	Different	4.04 (0.75) n=1210	-0.27 (0.99) n=194	0.01 (0.97) n=179	0.16 (0.90) n=245	0.04 (0.98) n=199	-0.02 (0.96) n=378	-0.03 (0.96) n=439	0.00 (0.97) n=756
	Same	4.12 (0.69) n=1770	-0.01 (0.97) n=279	0.18 (0.86) n=259	0.12 (0.90) n=383	0.06 (0.96) n=311	0.05 (0.94) n=529	0.07 (0.93) n=662	0.09 (0.93) n=1099
Maths	Different	3.79 (1.02) n=1210	0.09 (0.99) n=197	0.20 (0.93) n=182	-0.23 (1.04) n=246	-0.23 (1.06) n=201	0.05 (1.00) n=384	-0.09 (1.03) n=443	0.01 (1.01) n=767
	Same	3.86 (0.98) n=1769	0.17 (0.97) n=280	0.22 (0.94) n=264	-0.13 (0.98) n=382	-0.08 (1.01) n=313	0.09 (0.98) n=530	0.00 (0.98) n=662	0.07 (0.98) n=1107
Academic	Different	3.85 (0.62) n=1204	-0.03 (0.99) n=197	0.13 (1.01) n=176	-0.04 (0.98) n=246	-0.15 (1.04) n=198	-0.04 (1.01) n=380	-0.03 (0.98) n=443	-0.03 (1.02) n=754
	Same	3.91 (0.57) n=1753	0.07 (0.97) n=271	0.21 (0.96) n=260	-0.02 (0.94) n=381	-0.05 (0.94) n=308	0.12 (0.91) n=528	0.01 (0.95) n=652	0.09 (0.93) n=1096

Table 6.17. Self-perceived ability: Means, standard deviations (SD) and N for Quebec twin pairs taught by the same or different teachers at age 10, for the whole sample; by sex and zygosity; and by zygosity

Construct	Teacher	Whole sample	MZm	DZm	MZf	DZf	DZos	MZ	DZ
Reading	Different	3.91 (0.96) n=209	0.00 (1.04) n=56	-0.17 (1.19) n=37	-0.15 (1.03) n=52	-0.03 (1.02) n=43	0.13 (0.90) n=72	-0.07 (1.03) n=108	0.12 (1.01) n=152
	Same	3.91 (0.97) n=86	-0.10 (0.82) n=21	0.00 (1.05) n=16	0.12 (1.02) n=31	0.26 (0.77) n=16	-0.08 (1.12) n=25	0.03 (0.94) n=52	0.04 (1.00) n=57
Math	Different	3.70 (1.19) n=209	0.21 (1.03) n=56	-0.03 (1.04) n=37	-0.03 (0.92) n=52	-0.13 (1.07) n=43	0.15 (0.98) n=72	0.10 (0.98) n=108	0.02 (1.02) n=152
	Same	3.59 (1.22) n=86	-0.01 (0.80) n=21	0.31 (0.92) n=16	-0.21 (1.13) n=31	-0.08 (0.99) n=16	0.08 (1.00) n=25	-0.13 (1.00) n=52	0.09 (0.97) n=57

Table 6.16 & 6.17: Means for whole sample are raw and include outliers; means for zygosity and sex and zygosity groups are regressed on age with outliers removed. MZm = monozygotic male; DZm = dizygotic male; MZf = monozygotic female; DZf = dizygotic female; DZos = dizygotic opposite sex; MZ = all monozygotic; DZ = all dizygotic. Significant results in bold at $p \leq .001$ (.05/76).

Table 6.18. Self-perceived ability: Means, standard deviations (SD) and N for Quebec twin pairs taught by the same or different teachers at age 12, for the whole sample; by sex and zygosity; and by zygosity

Construct	Teacher	Whole sample	MZm	DZm	MZf	DZf	DZos	MZ	DZ
Reading	Different	3.67 (1.06) n=215	0.14 (0.89) n=36	-0.20 (1.09) n=33	-0.05 (0.84) n=48	-0.04 (1.00) n=36	-0.04 (1.03) n=65	0.03 (0.86) n=84	-0.08 (1.03) n=134
	Same	3.89 (1.05) n=143	-0.18 (1.14) n=27	0.10 (0.99) n=24	0.18 (0.95) n=35	0.15 (0.86) n=19	0.33 (0.87) n=38	0.02 (1.04) n=62	0.22 (0.90) n=81
Math	Different	3.57 (1.16) n=215	0.18 (1.04) n=36	0.00 (0.96) n=33	-0.10 (0.84) n=48	-0.60 (1.12) n=36	0.06 (0.98) n=65	0.02 (0.94) n=84	-0.19 (1.04) n=134
	Same	3.90 (1.10) n=143	0.31 (0.94) n=27	0.16 (1.07) n=24	0.11 (0.93) n=35	0.18 (0.87) n=19	0.15 (1.00) n=37	0.20 (0.93) n=62	0.16 (0.98) n=80

Means for whole sample are raw and include outliers; means for zygosity and sex and zygosity groups are regressed on age with outliers removed. MZm = monozygotic male; DZm = dizygotic male; MZf = monozygotic female; DZf = dizygotic female; DZos = dizygotic opposite sex; MZ = all monozygotic; DZ = all dizygotic. Significant results in bold at $p \leq .001$ (.05/76).

Table 6.19. Self-perceived ability: Means, standard deviations (SD) and N for the UK twin pairs taught by the same or different teachers at age 12, for the whole sample; by sex and zygosity; and by zygosity

Construct	Teacher	Whole sample	MZm	DZm	MZf	DZf	DZos	MZ	DZ
English	Different	3.91 (0.74) n=3471	-0.14 (0.95) n=465	-0.10 (0.97) n=447	0.10 (0.95) n=577	0.12 (0.94) n=493	0.02 (0.98) n=1007	-0.01 (0.96) n=1042	0.02 (0.97) n=1947
	Same	3.97 (0.71) n=1921	0.01 (0.98) n=265	-0.01 (0.90) n=246	0.19 (0.88) n=394	0.14 (0.87) n=339	-0.01 (0.96) n=464	0.11 (0.92) n=659	0.04 (0.92) n=1049
Maths	Different	3.79 (0.92) n=3472	0.16 (0.99) n=465	0.12 (0.98) n=448	-0.16 (0.94) n=583	-0.12 (0.95) n=490	0.01 (0.98) n=1007	-0.02 (0.98) n=1048	0.00 (0.98) n=1945
	Same	3.86 (0.89) n=1921	0.24 (0.94) n=266	0.19 (0.94) n=246	-0.08 (0.93) n=393	-0.01 (0.92) n=339	0.08 (0.97) n=466	0.05 (0.95) n=659	0.07 (0.95) n=1051
Academic	Different	3.87 (0.58) n=3472	0.02 (0.93) n=467	0.02 (0.96) n=449	-0.07 (0.96) n=578	-0.03 (1.01) n=493	0.02 (0.96) n=1006	-0.03 (0.95) n=1045	0.01 (0.97) n=1948
	Same	3.92 (0.56) n=1922	0.12 (0.98) n=266	0.12 (0.86) n=246	0.08 (0.93) n=392	0.04 (0.91) n=340	0.04 (0.93) n=465	0.10 (0.95) n=658	0.06 (0.91) n=1051
Academic Motivation	Different	**3.66 (0.57) n=3471**	-0.02 (0.94) n=467	-0.06 (0.96) n=447	-0.04 (0.98) n=579	-0.02 (0.99) n=493	0.02 (0.98) n=1011	-0.03 (0.96) n=1046	-0.01 (0.98) n=1951
	Same	**3.75 (0.56) n=1922**	0.06 (0.94) n=263	0.09 (0.93) n=247	0.10 (0.96) n=395	0.14 (0.94) n=339	0.03 (0.96) n=467	0.09 (0.95) n=658	0.08 (0.95) n=1053

Means for whole sample are raw and include outliers; means for zygosity and sex and zygosity groups are regressed on age with outliers removed. MZm = monozygotic male; DZm = dizygotic male; MZf = monozygotic female; DZf = dizygotic female; DZos = dizygotic opposite sex; MZ = all monozygotic; DZ = all dizygotic. Significant results in bold at $p \leq .001$ (.05/76).

Table 6.20. Teacher-student relations: Means, standard deviations (SD) and N for Quebec twin pairs taught by the same or different teachers at ages 7, 9, 10, and 12, for the whole sample; by sex and zygosity; and by zygosity

Age	Teacher	Whole sample	MZm	DZm	MZf	DZf	DZos	MZ	DZ
7	Different	3.91 (0.44) n=318	-0.31 (1.03) n=62	-0.43 (1.19) n=53	0.31 (0.76) n=65	0.20 (0.82) n=42	-0.05 (1.06) n=90	0.01 (0.95) n=127	-0.10 (1.08) n=185
	Same	3.94 (0.44) n=100	-0.19 (1.12) n=21	-0.22 (1.22) n=14	0.33 (0.81) n=27	0.11 (1.16) n=13	-0.03 (1.00) n=25	0.10 (0.98) n=48	-0.05 (1.09) n=52
9	Different	3.90 (0.43) n=291	-0.38 (0.97) n=57	-0.33 (1.07) n=43	0.36 (0.80) n=65	0.29 (0.83) n=40	-0.04 (1.02) n=84	0.02 (0.96) n=122	-0.04 (1.01) n=167
	Same	3.91 (0.46) n=88	-0.31 (1.18) n=16	-0.23 (0.93) n=14	0.35 (0.86) n=21	0.41 (0.97) n=13	-0.13 (1.17) n=24	0.06 (1.05) n=37	-0.02 (1.07) n=51
10	Different	3.87 (0.42) n=276	-0.34 (1.02) n=58	-0.50 (1.02) n=43	0.30 (0.81) n=53	0.13 (0.80) n=43	0.01 (0.90) n=76	-0.04 (0.98) n=111	-0.09 (0.94) n=162
	Same	3.93 (0.51) n=116	-0.03 (1.02) n=23	0.02 (1.30) n=16	0.22 (1.06) n=31	0.09 (1.23) n=17	0.03 (1.39) n=28	0.11 (1.04) n=54	0.04 (1.31) n=61
12	Different	3.88 (0.47) n=191	-0.38 (0.94) n=30	-0.51 (1.16) n=27	0.20 (0.94) n=40	0.41 (0.91) n=31	-0.09 (1.18) n=59	-0.05 (0.97) n=70	-0.05 (1.15) n=116
	Same	3.94 (0.44) n=125	-0.45 (1.08) n=19	0.24 (0.79) n=19	0.29 (0.77) n=31	-0.04 (1.32) n=17	0.02 (1.03) n=38	0.01 (0.96) n=50	0.06 (1.04) n=74

Means for whole sample are raw and include outliers; means for zygosity and sex and zygosity groups are regressed on age with outliers removed. MZm = monozygotic male; DZm = dizygotic male; MZf = monozygotic female; DZf = dizygotic female; DZos = dizygotic opposite sex; MZ = all monozygotic; DZ = all dizygotic. Significant results in bold at $p \leq .001$ (.05/76).

Table 6.21. Difference scores in school achievement: Means, standard deviations (SD) and N for difference scores between twin pairs taught by the same or different teachers for Quebec twin pairs at age 12, for the whole sample; by sex and zygosity; and by zygosity

Construct	Teacher	Whole sample	MZm	DZm	MZf	DZf	DZos	MZ	DZ
Reading	Different	0.79 (0.65) n=151	-0.08 (0.69) n=23	0.23 (1.04) n=20	-0.27 (0.69) n=32	-0.10 (0.88) n=25	0.43 (1.12) n=49	-0.19 (0.69) n=55	0.30 (1.05) n=94
	Same	0.59 (0.65) n=112	-0.37 (0.65) n=18	-0.03 (1.20) n=18	-0.40 (0.66) n=29	-0.38 (0.60) n=14	0.00 (1.17) n=32	-0.39 (0.65) n=47	-0.09 (1.08) n=64
Writing	Different	0.83 (0.78) n=153	-0.33 (0.86) n=23	0.37 (1.05) n=20	-0.26 (0.65) n=33	0.11 (0.95) n=25	0.34 (0.99) n=48	-0.29 (0.74) n=56	0.28 (0.99) n=93
	Same	0.61 (0.71) n=112	-0.11 (0.64) n=18	-0.07 (0.94) n=18	-0.47 (0.70) n=29	-0.26 (0.62) n=14	-0.27 (0.87) n=30	-0.33 (0.70) n=47	-0.21 (0.83) n=62
Maths	Different	0.76 (0.68) n=149	-0.30 (0.80) n=22	0.64 (1.14) n=20	-0.36 (0.69) n=32	-0.18 (0.93) n=24	0.43 (1.11) n=50	-0.33 (0.73) n=54	0.32 (1.10) n=94
	Same	0.63 (0.62) n=108	-0.57 (0.50) n=16	0.20 (0.87) n=18	-0.47 (0.74) n=27	0.28 (1.11) n=15	0.06 (1.08) n=32	-0.51 (0.65) n=43	0.15 (1.02) n=65
General	Different	0.73 (0.70) n=148	-0.09 (0.66) n=22	-0.06 (0.78) n=18	-0.37 (0.65) n=32	0.12 (0.76) n=25	0.43 (1.17) n=48	-0.25 (0.66) n=54	0.25 (1.01) n=91
	Same	0.55 (0.64) n=111	-0.56 (0.71) n=18	0.12 (0.89) n=18	-0.40 (0.65) n=29	-0.18 (0.77) n=14	0.05 (1.03) n=31	-0.46 (0.67) n=47	0.02 (0.93) n=63

Means for whole sample are raw and include outliers; means for zygosity and sex and zygosity groups are regressed on age with outliers removed. MZm = monozygotic male; DZm = dizygotic male; MZf = monozygotic female; DZf = dizygotic female; DZos = dizygotic opposite sex; MZ = all monozygotic; DZ = all dizygotic. Significant results in bold at $p \le .005$ (.05/11).

Table 6.22. Difference scores in school achievement: Means, standard deviations (SD) and N for difference scores between twin pairs taught by the same or different teachers for the UK twin pairs at age 16, for the whole sample; by sex and zygosity; and by zygosity

Construct	Teacher	Whole sample	MZm	DZm	MZf	DZf	DZos	MZ	DZ
Maths GCSE	Different	**0.66 (0.68) n=1142**	-0.34 (0.74) n=161	**0.24 (0.92) n=166**	**-0.29 (0.68) n=166**	**0.08 (0.92) n=215**	**0.17 (0.90) n=408**	**-0.31 (0.71) n=327**	**0.16 (0.91) n=789**
	Same	**0.32 (0.46) n=468**	-0.58 (0.48) n=88	**-0.30 (0.71) n=65**	**-0.53 (0.57) n=136**	**-0.30 (0.70) n=73**	**-0.32 (0.66) n=104**	**-0.55 (0.54) n=224**	**-0.31 (0.68) n=242**
English GCSE	Different	**0.66 (0.67) n=1219**	-0.18 (0.77) n=172	0.18 (1.01) n=190	**-0.28 (0.63) n=195**	**0.10 (0.92) n=204**	**0.17 (0.95) n=442**	**-0.24 (0.70) n=367**	**0.15 (0.96) n=836**
	Same	**0.37 (0.47) n=411**	-0.43 (0.62) n=82	-0.04 (0.83) n=50	**-0.55 (0.54) n=113**	**-0.31 (0.71) n=85**	**-0.20 (0.77) n=81**	**-0.50 (0.58) n=195**	**-0.20 (0.76) n=216**

Means for whole sample are raw and include outliers; means for zygosity and sex and zygosity groups are regressed on age with outliers removed. MZm = monozygotic male; DZm = dizygotic male; MZf = monozygotic female; DZf = dizygotic female; DZos = dizygotic opposite sex; MZ = all monozygotic; DZ = all dizygotic. Significant results in bold at $p \leq .005$ (.05/11).

Table 6.23. Difference scores in motivational constructs; and teacher-student (T/S) relation: Means, standard deviations (SD) and N for difference scores between twin pairs taught by the same or different teachers for Quebec twin pairs at age 12, for the whole sample; by sex and zygosity; and by zygosity

Construct	Teacher	Whole sample	MZm	DZm	MZf	DZf	DZos	MZ	DZ
Reading enjoyment	Different	0.95 (0.75) n=211	-0.28 (0.80) n=34	-0.01 (0.91) n=32	-0.21 (0.87) n=46	0.24 (1.00) n=34	0.40 (1.14) n=63	-0.24 (0.84) n=80	0.25 (1.06) n=129
	Same	0.81 (0.73) n=143	-0.36 (0.88) n=27	0.13 (1.22) n=24	-0.55 (0.65) n=35	-0.02 (1.13) n=19	0.26 (0.95) n=37	-0.47 (0.75) n=62	0.15 (1.07) n=80
Math enjoyment	Different	1.02 (0.76) n=211	-0.17 (0.99) n=34	0.22 (0.96) n=32	-0.10 (0.82) n=46	0.11 (0.91) n=33	0.20 (1.08) n=63	-0.13 (0.89) n=80	0.18 (1.00) n=128
	Same	0.89 (0.76) n=143	-0.35 (0.87) n=27	0.19 (1.26) n=24	-0.40 (0.70) n=34	0.23 (1.05) n=19	-0.07 (0.95) n=37	-0.38 (0.77) n=61	0.08 (1.07) n=80
Reading SPA	Different	1.02 (0.79) n=211	-0.03 (1.03) n=34	0.13 (0.88) n=32	-0.26 (0.75) n=46	0.14 (0.95) n=33	0.19 (1.04) n=63	-0.16 (0.88) n=80	0.16 (0.97) n=128
	Same	0.91 (0.80) n=143	-0.13 (0.99) n=27	-0.17 (0.87) n=22	-0.14 (0.98) n=35	0.04 (0.97) n=19	-0.15 (0.94) n=37	-0.13 (0.97) n=62	-0.11 (0.92) n=78
Math SPA	Different	1.01 (0.82) n=211	-0.45 (0.85) n=34	0.28 (1.13) n=32	-0.28 (0.75) n=46	0.29 (0.95) n=33	0.26 (1.05) n=63	-0.35 (0.79) n=80	0.27 (1.04) n=128
	Same	0.91 (0.78) n=143	-0.10 (1.05) n=27	0.09 (0.95) n=23	-0.34 (0.89) n=35	0.09 (0.90) n=19	-0.11 (0.91) n=37	-0.24 (0.96) n=62	0.00 (0.91) n=79
T/S relations	Different	0.40 (0.79) n=233	0.37 (1.28) n=37	0.38 (1.29) n=33	-0.01 (1.01) n=50	0.16 (1.15) n=36	-0.02 (0.98) n=67	0.15 (1.14) n=87	0.13 (1.12) n=135
	Same	0.19 (0.58) n=152	0.24 (1.21) n=27	-0.17 (0.82) n=24	-0.25 (0.67) n=36	-0.13 (0.88) n=19	-0.27 (0.62) n=42	-0.04 (0.96) n=63	-0.21 (0.73) n=85

Means for whole sample are raw and include outliers; means for zygosity and sex and zygosity groups are regressed on age with outliers removed. MZm = monozygotic male; DZm = dizygotic male; MZf = monozygotic female; DZf = dizygotic female; DZos = dizygotic opposite sex; MZ = all monozygotic; DZ = all dizygotic. Significant results in bold at $p \leq .005$ (.05/11).

Table 6.24. Percentage of twin pairs (by zygosity) taught by the same or different (S/D) teachers most of the time for Quebec and the UK samples at ages 9 and 10 years

Age	Country	S/D teacher all/most of the time	MZ	DZ	Total
Age 9	Quebec Canada	Different all years	62.4% n=58	62.4% n=73	62.4% n=131
		Same all years	4.3% n=14	7.7% n=13	6.2% n=27
		Different age 7 same age 9	4.3% n=4	7.7% n=9	6.2% n=13
		Same age 7 different age 9	18.3% n=17	18.8% n=22	18.6% n=39
		Total	100% n=93	100% n=117	100% N=210
	UK	Different all years	28.4% n=318	29.6% n=570	29.2% n=888
		Same all years	53.1% n=595	52.5% n=1011	52.7% n=1606
		Different age 7 same age 9	6.9% n=77	5.9% n=114	6.3% n=191
		Same age 7 different age 9	11.7% n=131	11.9% n=230	11.9% n=361
		Total	100% n=1121	100% n=1925	100% N=3046
Age 10	Quebec Canada	Different all years	55.8% n=43	65.9% n=56	61.1% n=99
		Same all years	16.9% n=13	8.2% n=7	12.3% n=20
		Same most years	3.9% n=3	3.5% n=3	3.7% n=6
		Different most years	20.8% n=16	16.5% n=14	18.5% n=30
		Equal number of same/different years	2.6% n=2	5.9% n=5	4.3% n=7
		Total	100% n=77	100% n=85	100% N=162
	UK	Different all years	28.7% n=261	30.9% n=483	30.1% n=744
		Same all years	51.0% n=463	49.4% n=771	50.0% n=1234
		Same most years	7.7% n=70	6.3% n=98	6.8% n=168
		Different most years	12.6% n=114	13.4% n=210	13.1% n=324
		Total	28.7% n=261	30.9% n=483	30.1% N=744

MZ = monozygotic twins; DZ = dizygotic twins.

Table 6.25. Percentage of twin pairs (by zygosity) taught by the same or different (S/D) teachers most of the time for Quebec and the UK samples at ages 12 and 14 years

Age	Country	S/D teacher all/most of the time	MZ	DZ	Total
Age 12	Quebec Canada	Different all years	51.9% n=27	61.9% n=39	57.4% n=66
		Same all years	15.4% n=8	7.9% n=5	11.3% n=13
		Same most years	1.9% n=1	4.8% n=3	3.5% n=4
		Different most years	30.8% n=16	25.4% n=16	27.8% n=32
		Total	100% n=52	100% n=63	100% N=115
	UK	Different all years	35.8% n=190	43.0% n=364	40.2% n=554
		Same all years	33.0% n=175	23.5% n=199	27.2% n=374
		Same most years	3.8% n=20	2.6% n=22	3.1% n=42
		Different most years	15.1% n=80	19.7% n=167	17.9% n=247
		Equal number of same/different years	12.3% n=65	11.2% n=95	11.6% n=160
		Total	100% n=530	100% n=847	100% N=1377
Age 14	UK	Different all years	45.0% n=148	49.3% n=255	47.6% n=403
		Same all years	16.4% n=54	11.2% n=58	13.2% n=112
		Same most years	5.8% n=19	4.1% n=21	4.7% n=40
		Different most years	32.8% n=108	35.4% n=183	34.4% n=291
		Total	100% n=329	100% n=517	100% N=846

MZ = monozygotic twins; DZ = dizygotic twins; Age 16 was excluded, as there was significant loss of power due to attrition

Table 6.26. Achievement and motivation: ANOVA results at age 12 (Quebec-Canada) and age 16 (UK) by zygosity and a cumulative effect of being taught by the same or different (S/D) teachers most of the time during years of education

Country	**School subject/test**	**S/D class mostly**		**Zygosity**		**S/D class mostly*zygosity**	
		p	**η^2**	***p***	**η^2**	***p***	**η^2**
Quebec Canada	Reading	.958^	.000^	.848	.000	.096^	.009^
	Writing	.955	.000	.396	.002	.101^	.009^
	Maths	.610	.001	.950	.000	.295^	.004^
	In General	.552^	.001^	.860	.000	.019^	.018^
	Reading enjoyment	.553	.001	.602	.001	.583^	.001^
	Maths enjoyment	.252	.003	.152	.005	.401	.002
	Reading SPA	.977	.000	.890	.000	.547	.001
	Maths SPA	.962	.000	.119	.006	.358	.002
	Teacher-student	.304^	.003^	.857	.000	.488^	.002^
UK	Maths GCSE	.391	.001	.291	.002	.277	.002
	English GCSE	.724^	.000^	.784^	.000^	.431^	.001^

Analyses at age 16 (UK) were conducted on twins taught by same or different teacher for most of their school years. Bold indicates significance with a Bonferroni multiple testing correction applied of $p = .05$ divided by the number of measures (k=11) across ages 12 and 16 and across both samples which provided a significance value of $p \leq .005$ (.05/11). ^ = unequal variance: Levene's test significant at $p \leq .05$.

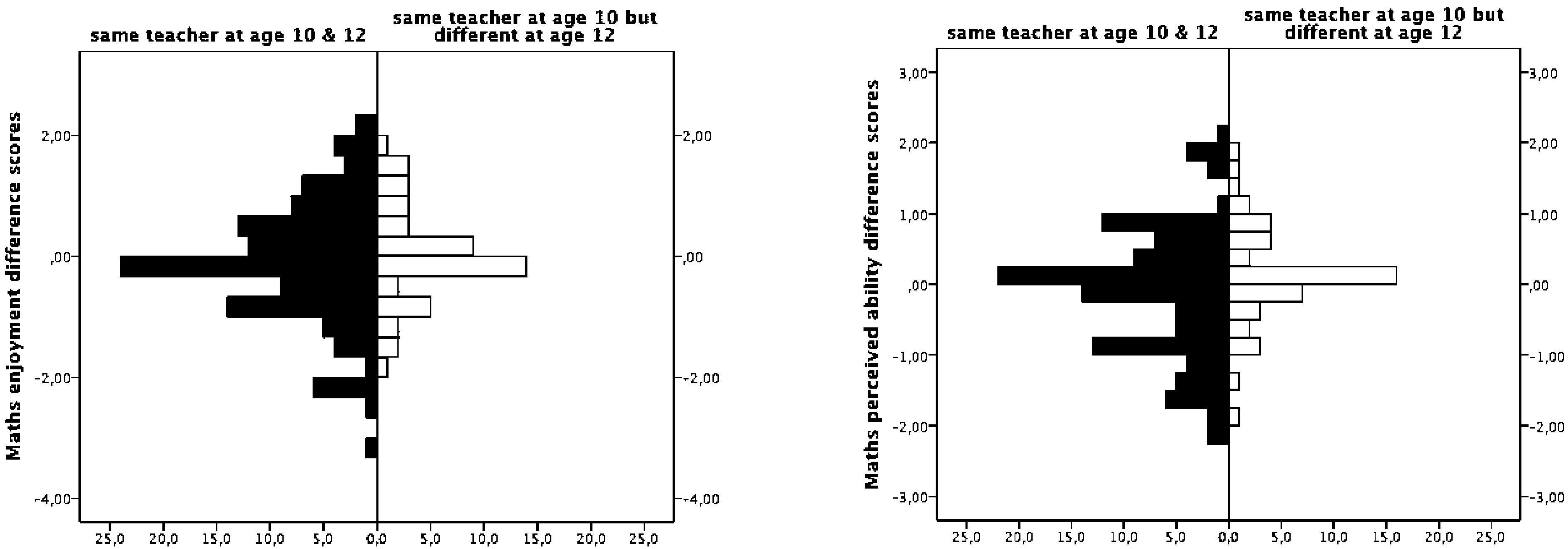

Figure 6.2. Difference scores in maths enjoyment and maths perceived ability for two groups of Quebec twin pairs: twins taught together at both age 10 and age 12; and twins taught together at age 10 and separately at age 12. Difference scores were calculated between twin pairs taught together across age 10 and 12, and between twin pairs taught together at age 10 then separately at age 12. Positive values on the y-axis indicate greater similarity between twin pairs, while negative values indicate greater difference between twins. The x-axis indicates the frequency for a specific y-axis value.

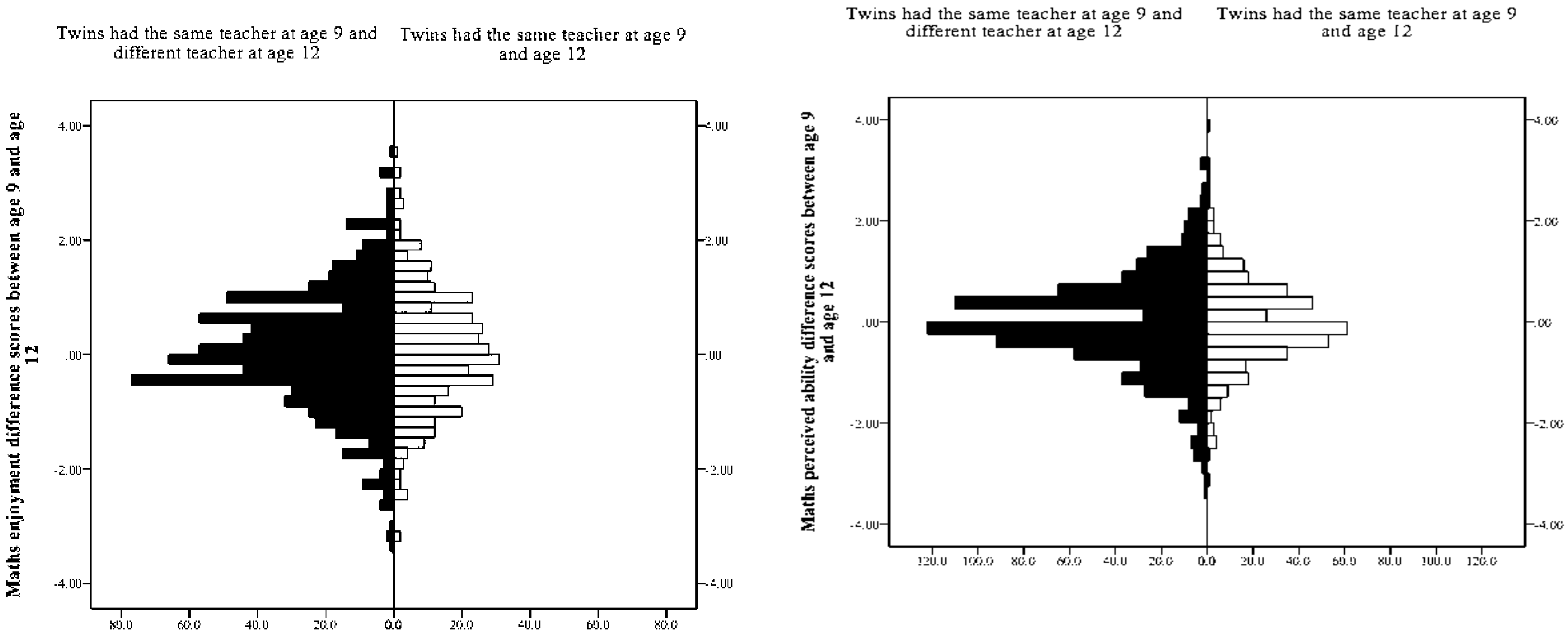

Figure 6.3. Difference scores in maths enjoyment and maths perceived ability for two groups of UK twin pairs: twins taught together at ages 9, 10 and age 12; and twins taught together at ages 9 and 10 but separately at age 12. Difference scores were calculated between twin pairs taught together across age 9 and 12, and between twin pairs taught together at age 9 then separately at age 12. Positive values on the y-axis indicate greater similarity between twin pairs, while negative values indicate greater difference between twins. The x-axis indicates the frequency for a specific y-axis value.