Assessment: feedback from our pasts, feedforward for our futures

Kay Stables

Introduction

Some years ago, on behalf of *Design and Technology: An International Journal*, I guest-edited an issue of the journal on the topic of assessment. In introducing the theme of the special edition, I drew an example of learning from my personal life that highlighted some important issues about assessment, and I'm taking the liberty of repeating the example here as a starting point from which to explore the current state of play in assessment and considerations for how future agendas might be developed. The example reflects on my experience of buying a new washing machine.

> *One of the realities of learning processes is that you only know that, or how well, you have learned something if there is some kind of feedback in the system. I know that I have learnt how to make my new washing machine work when I put in dirty clothes, and some time later take out clean ones. I also get a whole series of clues about my success along the way: I turn a dial and a light indicates that I have selected a particular wash programme; I re-set the dial and find that I can choose between more than one programme; a further light indicates which button has to be pressed to start the process— and when the button doesn't work another one indicates that I have yet to close the door. Then comes the reassuring sound of water flooding into the machine ... I have succeeded in getting the process off the ground. The feedback I am receiving is doing two things. It is reassuring me when I get things right and it is providing clues or prompts to help me learn. If I get really stuck, there is always the instruction book to refer to, which gives more detailed and authoritative advice, although often far more than I need at that particular moment.*
>
> *So, with good feedback, a sprinkling of problem solving, a level of personal confidence and an authoritative source to turn to when needed, my learning in respect of using a washing machine makes good progress. What is notable though, is that before I engaged in this process, no one assessed me to see if I was capable of using a washing machine and at the end no one assessed me to see what standard in using washing machines I had reached. And yet the learning took place.*
> (Stables, 2007, p.3)

What might be seen as the implication of this example is that we don't need assessment to learn. If this is the case, why dedicate a whole book chapter to the topic? But the example is set in a very particular context—the learner had real motivation to learn—a genuine 'need to know' in order to get the washing clean. The learner was also a confident adult with considerable life experience (including of washing machines) to draw on, already had a repertoire of approaches to problem solving and self discovery, and was receiving feedback 'prompts' from the machine itself. In many learning situations things are not quite so benign. And in many assessment situations, supporting learning is not a priority. The name of the game is labeling and sorting: can use a washing machine; can't use a washing machine. For some, 'good' assessment is that which is supporting the development of a learner. For others, 'good' assessment is that which allows us to reliably categorise people by what they can or can't do, or know or don't know. These different standpoints can be witnessed in education systems and settings across the globe and can be extremely problematic when trying to agree on valuable, effective systems of assessment. So how did we get to where we are?

It has been recorded that formal written tests go back over 2,000 years. Introduced in China, written tests were created to provide a meritocratic route into becoming a civil servant—"for

identifying the talented among the common people" (Hanson, 1994, p.186). The history of these very early assessments is fascinating (see Hanson, 1994) but for our purpose it is interesting to note, briefly, the age-old assessment issues that the history displays. Parental involvement in preparing young people for the test even went so far as pregnant mothers being exposed to test-relevant content. Young people 'crammed' for the tests by being locked in a room for a year. So many took the tests that they were locked in cells for three days to complete them, first having been rigorously searched to avoid cheating and checked to see that the person wasn't an imposter. Extreme lengths were employed to ensure candidates' submissions were anonymous. The tests themselves were so focused on standardised results based on correct responses that, despite their meritocratic underpinning of seeking talent in 'common people', this standardisation was what finally brought the system down. It stifled creativity and resulted in an assessment system that didn't provide the evidence that was required—to know whether the person tested was qualified to become an effective civil servant.

Whilst being quite extreme, it is surprising how the spirit of this early approach still haunts us today. Kelly (2009) cites the Taunton Report (1868) and the Beloe Report (SSEC, 1960) in reminding us that "[t]he assessment (and evaluation) tail will always wag the curriculum dog" (p. 148). We have also seen in Technology Education how assessment has distorted the curriculum (e.g., Atkinson 2000; Harris & Wilson, 2003; Kimbell, 1997, 2006; OfSted 2001, 2002). Even the overarching aim of a curriculum can be undermined, as was seen in the Prest Review (2002) where, for example, an aim of the English National Curriculum for Design and Technology (D&T) was for learners to "think and intervene creatively" but the assessment criteria made no mention of creativity, resulting in it being undervalued and under-prioritised in classroom activities.

Different purposes, different approaches

Assessment of learning or assessment for learning?

Much of the history of assessment is based in what Hanson (1994) refers to as "qualifying tests" that exist to see if a person is fit to move into a particular place within society and also within a particular education system. In education, the history has been of the development of qualifying tests as 'entrance exams', for example, to gain a place in University. These entrance exams soon spawned the development of 'exit' exams, such as school leaving exams, which created a chain of 'qualifying tests': take an entrance exam to see if you qualify for a school place, take an exit exam at school to see if you qualify for a university place, and so on. Such assessment systems are ubiquitous in school systems across the globe and across disciplines, including Technology Education. Described by MacLeod (1982) as "the single most intrusive and expensive innovation in Western education", they are often accompanied by disturbing effects, as Hanson describes:

> regardless of people's feelings about them, qualifying tests are a key factor for living successfully in contemporary society. Those who reject the message of personal insufficiency reiterated by poor test performance may turn off on tests, but then the system turns off on them. They are excluded from educational opportunities and good jobs and (just as the tests predicted!) they never are able to accomplish much. ... Qualifying tests constitute one of the central conditions of contemporary society. (p. 186)

The contrast between my *personalised* learning in the washing machine example and Hanson's analysis of effects of a history of qualifying tests could not be more stark and yet they can both be viewed on an assessment continuum as, for example, characterised by Harlen & Deakin Crick (2002): "Assessment is a term that covers any activity in which evidence of learning is collected in a planned and systematic way, and is used to make a judgment about learning" (p. 1).

The term assessment has now become so common and embedded in our education systems that there is a danger that a single view is assumed of what it is taken to mean. The reality is that the term has many meanings, behind which lie a plethora of different functions of assessment and different philosophies of learning.

2

Much has been written about the purposes of assessment. A useful and straightforward lens through which to view these is provided through the extension of Harlen and Deakin Crick's (2002) definition used above, in which they dichotomise the function of assessment:

> If the purpose is to help in decisions about how to advance learning and the judgement is about the next steps in learning and how to take them, then the assessment is formative in function. If the purpose is to summarise the learning that had taken place in order to grade, certificate or record progress, then the assessment is summative in function. (p. 1)

Assessment purposes can therefore be seen on a continuum. At one end sit the formal 'qualifying tests' described by Hanson—assessment *of* learning. At the other end sits my washing machine example—assessment *for* learning, assessment to support the learning as it is taking place.

Written or practical?

When we turn to a curriculum area such as Technology Education, the territory of assessment becomes further complicated by the practice-based nature of the subject. The examples of qualifying tests alluded to earlier are largely of written tests. Yet early examples of qualifying tests of practice also exist, for example, those initiated in mediaeval times that were the basis of qualifying within the guild of a particular craft through an apprenticeship model of learning. Assessment in Technology Education has developed from both roots—the qualifying written test and the qualifying practical test. But the roots of each grow in different soil, the written tests coming from an academic tradition where 'knowledge' can be demonstrated through the written word, the practice-based tests coming from what Fleming (2013) describes as a holistic model where learning is through mentoring and direct experience and where 'knowledge' is embedded in the quality of material artefacts. These dual systems have presented conflict and challenges for Technology Education, exacerbated by the seemingly higher status of the written test and its link to an academic education and the seemingly lower status of the practical test and its link to vocational education. To bring further understanding to the complexities in assessment, we turn next to who assessment is for—who are the stakeholders in the process?

Stakeholders in assessment

The range of stakeholders in educational assessment systems is diverse, including learners, teachers, parents and guardians, school administrators, policy makers, politicians and employers. With each stakeholder group come different sets of understandings, values and agendas. The world of assessment can be presented as being as simple and straightforward as Harlen and Deakin Crick's clear categorisation might appear to suggest, but the reality is likely to be far more complex and messy.

Taking into account the stakeholders in any assessment situation adds a further layer to this complexity, increased still further when we consider how different stakeholders perceive feedback from assessments, and the potential consequent actions (the feedforwards) that are conceived. If the teacher is the stakeholder in formative assessment, then the purpose may become diagnostic, as they gain better insight into the learner's understanding, ability, skill, etc. and plan the next step in learning. If the learner is the stakeholder, the purpose may be metacognitive, as the learner comes to have a better understanding of her or his own understanding and how this can be used in future situations. Summative assessment for a teacher may be evaluative as he or she reflects on the learning experiences that have been provided and how these experiences might be modified in the future. For a policy maker however, the feedback from summative assessments may be read quite differently, for example, resulting in decisions on how the teaching workforce is paid.

Historically there has been a dominating thrust, particularly in Western education, for assessment to be focused on the summative function of producing grades and certification as evidence of 'exit' qualifications to inform the gatekeepers of future education or employment. In prioritising in

this way, assessments create extrinsic motivation for the learner under scrutiny. However, recent decades have seen a swing of the pendulum to more pedagogic approaches to assessment where the primary focus has been to support the processes of learning. An example of this swing can be seen in policy documentation that supported the initial development of a National Curriculum (NC) in England and Wales in 1990. From the report of the Task Group on Assessment and Testing (the TGAT report, DES, 1988) led by Professor Paul Black, we have a clear statement of this shift:

> Promoting children's learning is a principal aim of schools. Assessment lies at the heart of this process. It can provide a framework in which educational objectives may be set, and pupils' progress charted and expressed. It can yield a basis for planning the next educational steps in response to children's needs. (para 3)

This statement illustrates a paradigm shift to a pedagogic approach, and similar shifts can be witnessed in other curriculum settings. However, the reality of pendula is that they keep on swinging and, in the context of educational assessment, often create unhelpful shifts and tensions in policy that have challenging impacts in classrooms for both teachers and learners.

The pendulum swing in the context the English NC has been a turbulent tug-of-war between the educationalists and politicians, policy makers and industrial stakeholders. The former have brought a focus on the learner to centre stage while the latter groups have brought priorities such as the economy to the forefront. Technology Education can be caught in the crossfire of such battles as, for example, has been seen in England where the GCSE (16+ exit examination) has seen the balance between continuous assessment of coursework (that suits the practice-based ethos of D&T) and end of course summative assessments swing dramatically. From 60% coursework when GCSEs were introduced in 1988, the first shift was in 2009 to a maximum 60% 'controlled assessment' (i.e., coursework conducted under exam conditions so that there could be no 'cheating' or 'parental help') to a current (2013) proposal to remove all coursework from GCSEs, with the exception of practical subjects (including D&T) where an amount determined by strict principles may be allowed (Ofqual, 2013).

Contested and conflicting philosophies

Underpinning conflicting positions of different stakeholders is the age-old dichotomy of what we are educating learners for. Where school education is seen to be important in the context of the rounded education of the whole individual for life, there is typically an emphasis on intrinsic drivers for learning and, following from this, attainment and achievement. Where a more instrumental view is taken, assessment tends to be more focused on extrinsic drivers of attaining the right grades and, particularly in exit assessments, towards fitting into the workplace of a society.

When the economic position of a society is under threat, policy makers turn to education as a way of solving this problem. This can be seen throughout history as industrialisation and—more recently globalisation—have made their mark, the result of which has been to politicise assessment and increase the emphasis on more instrumental goals. This shift can be seen in the increased attention that governments pay not just to national comparisons of achievement but to where they stand on the international stage as identified through assessments such as the PISA tests. The emphasis on instrumentally-focused assessment encourages what might be seen as the more traditional positivist approach to assessment based on predetermined outcomes and performance criteria that "takes away from the originality, criticality and creativity of the work" (Elton, 2006, p.124).

Once assessment becomes politically loaded, systems are thrown into tension between the focus on the individual or on whole populations and on local agendas or global agendas. While educators are more likely to favour individual approaches that focus on culturally and socially relevant assessment, politicians concerned with their rankings in international league tables are

more likely to focus on generalised and traditional knowledge-focused assessments that match the positivist views described above by Elton.

Dipping into examples highlights the challenges that these tensions bring. Pellegrino (2006), analysing the problems of a "flawed and broken" assessment in the USA, sees a system in need of radical change,

> so that it can support processes of teaching and learning focused on deep learning and understanding … the dollars we now spend on an assessment should be reinvested in more targeted and efficacious assessment approaches tied to important curricular goals. These assessments should be meaningful to the individuals assessed and have real value in determining their readiness to move on in the educational system. (Pellegrino, 2006, p. 2)

Working alongside other educationalists and also industrialists, Pellegrino contributed to the National Centre on Education and the Economy's (NCEE) *New Commission on the Skills of the American Workforce* to seek a solution to this 'broken' system. The report, laying out the territory for a new approach (NCEE, 2007), makes it clear that the Commission's concern is with the steep decline in the United States' place in international league tables of educational attainment. Reflecting back to the report by the first Commission in 1990, the authors comment that

> [t]he first Commission never dreamed that we would end up competing with countries that could offer large numbers of highly educated workers willing to work for low wages. But China and India are doing exactly that. Indeed, it turns out that China and India are only the tip of the iceberg. .... Thirty years ago, the United States could lay claim to having 30 percent of the world's population of college students. Today that proportion has fallen to 14 percent and is continuing to fall. (p. 4)

Where assessment is concerned, the report calls for development of "standards, assessments, and curriculum that reflect today's needs and tomorrow's requirements" and criticizes current school exit exams that "measure the acquisition of discipline-based knowledge in the core subjects", wishing to see these replaced with qualities they perceive to be needed in the 21st Century, such as creativity and innovation, ideas and abstractions, self-discipline, function in a team, etc. This report has led to the recent development of "Excellence for all", a pilot programme that has created an alternative curriculum and assessment system for high school students designed to lead them to higher education or skilled employment. The system focuses on the new 'Core Curriculum' that excludes Technology Education.

The issue of vocationalism is also a very real one for Technology Educators—where the balance between educating for life and educating for a technological job can bring different priorities both in curriculum and in assessment, resulting in summative assessments that focus less on a holistic view of capability within a practice-based discipline and more on specific and isolated vocationally-related knowledge and skills. This split view is compounded by different paradigms operating with curriculum and assessment in what Shepard (2000) identifies as a disjuncture between assessment systems still operating on a behaviourist paradigm while curriculum and 'instruction' have moved towards new paradigms of constructivist and sociocultural learning (see David Mioduser's chapter, this volume).

These alternative views of learning can be seen as underpinning some of the contradictions—that even when teachers have progressed approaches to learning and teaching, they still operate within a non-aligned model of traditional assessment. Furthermore, even when policy has moved forward, teachers may still cling to traditional practices. Of course, this issue is not unique to the USA. For example, it has been noted by Beets and van Louw (2011) in the context of policy developments in South Africa, where they found that policy is ahead of many teachers whose teacher-directed approaches create an ethos that suggests assessment is something that teaches do to learners, not that learners are engaged in for themselves.

The impact of assessment systems on learners and learning

Few people enjoy being tested and for some the level of stress is entirely counterproductive in relation to learning. This is exacerbated when assessment is divorced from the processes of learning. Harlen and Deakin Crick (2002) explored the impact of high stakes assessment policy on learner motivation. Analysing existing research from a range of countries (Canada, Israel, Morocco, Northern Ireland, UK and USA) on the impact of summative assessment on motivation, they found that where there was impact, it was largely negative. They reported that low achieving learners experienced a lowering of self esteem when faced with summative assessment, that learners (and especially girls) suffered anxiety, and that the ethos in classrooms changed such that learners perceived all assessments as summative even when the teachers intended them as formative. Teaching styles became more transmissional (echoing the disjuncture of paradigms highlighted by Shepard, 2000), which created a bias towards learners who responded well to this style. The emphasis on learning shifted from process to performance and students felt the pressure of high stakes assessment whether the consequences were personal (as in the eleven plus exams in Northern Ireland) or more school focused (as in England's NC assessments or SAT assessments in the United States). They drew important messages for both practice and policy from their study, suggesting a shift in practice to focus process over performance, to develop a constructive and supportive ethos to assessment, cultivate intrinsic motivation and self assessment and for policy makers to additionally consider the issues the study raised about the validity of assessments that had such negative and cumulative impacts on learners. This research indicates important issues for Technology Education. As a process-based learning area where creativity and risk-taking are vital, learning needs to be enacted in environments that nurture self-esteem and motivation if learners are to develop confidence and competence as technologists and designers.

A further set of issues focusing on equity are presented by Beets and van Louw (2011) through an analysis of the impact of policy on assessment practices in the context of post-Apartheid South Africa. With reference to Bourdieu's concept of cultural capital, they draw attention to the extent to which assessments are influenced by the way in which educational resources are distributed and that assessments are aimed at a societal norm,

> aimed at able-bodied learners, who are in command of the (instructional) languages of the assessors, who are appropriately prepared (had access to good teaching, etc.) and who are able to read, write and understand what is presented to them. (p.311)

For South Africa, as with many countries where multiple indigenous languages sit alongside a mainstream language, the language of assessments is critical in ensuring validity and reliability. In the South African context, learning and assessments are carried out for the majority of children in a second or even a third language. Beets and van Louw draw attention to the extent to which the learner's own language, a key aspect of their cultural capital, is being marginalised through the medium of assessment, with consequent impediment on individual's ability to achieve.

Each of the examples above illustrates the tensions that arise between the assessment policies that governments create and the practices that are enacted in classrooms. Each also draws attention to the impacts on learning and teaching that these tensions create. What is abundantly clear is that any assessment system is likely to have unintended consequences. But the more these consequences are brought to the fore, the greater the chance that any new systems can take into account and minimise the untoward.

Assessment structures and practices

Whilst it is useful to understand the drive behind instrumental views of assessment, educators are more likely to see value in approaches that are designed to support the development of the learner and by inference the learning experiences that teachers provide. Viewed from the standpoints of reliability and validity, positivist, instrumental approaches are likely to be more concerned with reliability and interpretivist, liberal approaches more concerned with validity. As government agendas are often the dominant driver in educational assessment, the educationists

are often put in the position of seeking the 'added value' of validity to get the best out of a system on behalf of the learners. This position could be seen as one of compromise and I would argue that we should (always) be seeking to optimise a system. I would also argue that future-facing assessment structures and practices in Technology Education should prioritise validity concerns and then aim to gain added value from achieving reliability.

This view accords well with the concept of sustainable assessment developed by David Boud (2000, 2007) through which, in reference to Higher Education assessment practices, he proposes

> *assessment practices that met the needs of an institution to certify or provide feedback on students' work, but which would also meet the long-term need of equipping students for a lifetime of learning (Boud, 2000). In this way of viewing assessment, every act of assessment would do double duty in addressing the immediate needs of the institution while also contributing in some way to the development of the skills and capabilities of students to be effective learners. (Boud, 2007, p7)*

Anning, Cullen and Fleer (2009), working from a sociocultural-historical perspective in Early Childhood Education make a similar point:

> In working towards the future, early childhood teachers need two types of conceptual tool. The first tool is built upon socio-cultural theory, where documentation of learning moves beyond an individualistic orientation and acknowledges that learning is owned by a community of learners. In building learning stories and in mapping the transformation of understanding greater insights can be gained about children's learning and teachers' teaching. Secondly, the profession needs instruments which can extract from this rich web of assessment activity discrete measures of understanding as matched to government priorities. (Anning, Cullen, Fleer, 2009 p.194)

Identifying an aspiration to achieve this optimisation is one thing. Finding the practices that support the aspiration is something else. Critical in doing this is, I believe, the need to identify some underlying principles that can help all stakeholders, and most importantly teachers, maintain a balance between the different drivers of assessment policies and practices and take ownership of how these impact on learners and learning.

The issues raised so far in this chapter, including the need to shift the dominant behaviourist assessment paradigm to align with more social constructivist learning (Shepard, 2000) provide some indications that can be drawn on in identifying such principles. Research in D&T education that we have conducted over nearly thirty years in the Technology Education Research Unit (TERU) complements and grounds the above and it is to this research that I now turn. On the basis of both the issues raised and on grounded research I propose the following are key in creating guiding principles:

- Adopting a pedagogic approach to assessment
- Maintaining authenticity in activities through which assessment is being undertaken
- Recognising the importance of judgement in valid processes of assessment
- Maintaining a focus on equity and the inclusive role of the learner.

A pedagogic approach of assessment

If we believe that assessment is fundamental in learning, then is there any learning activity that, potentially, can't be optimised by also being an assessment activity? Likewise, is there any assessment activity that can't be optimised in terms of how it also supports learning? While there are clearly plenty of examples of both teaching and assessment where this is not the case, in an ideal world why would we not want it to be so? Where teaching and assessment practices explicitly aim to support learning, we have what can be termed a pedagogical model of assessment. In many ways such a model can be seen as an efficient one, particularly in terms of the time saved from the preparation for, and taking of, more traditional assessments.

7

This approach was one that we took when creating Technology Standard Assessment Tasks for five to seven year olds at the introduction of a Technology NC in England and Wales in 1990. The approach was important. Not only were teachers not experienced in the criterion and levels-based assessment that the NC introduced, but it was the first time that Early Years teachers were teaching a subject formally called Technology. For many, teaching what they saw as a new subject was daunting enough, let alone having to assess the learners as well. Consequently we focused on the development of good learning and teaching activities, designed to scaffold teachers' understanding of structuring and implementing Technology projects that would also provide evidence of learning. The activities were accompanied by exemplified guidance for assessment against the NC criteria (Stables, 1992). Effectively we were supporting the development of the teachers' Pedagogic Content Knowledge (PCK), although this was not the language used at the time. The sets of SATs that we generated were designed to model learning, teaching and assessment in Early Years Technology activities. The approach was one that saw learning and assessment as symbiotic. The assessment activities were structured to work as stand-alone learning and teaching activities, and yet also as formative, diagnostic and summative assessment activities. In developing them we were mindful of the relationship between the learning intention, the assessment intention and the activity designed to support both—what Biggs (1996) terms "constructive alignment" (p. 347.

More recently we have likened this process to seeing this activity through a mirror where both teacher and learner can see double-sided reflections supporting summative and formative assessment and also learning and teaching (Kimbell & Stables, 2008b), as shown in Figure 1.
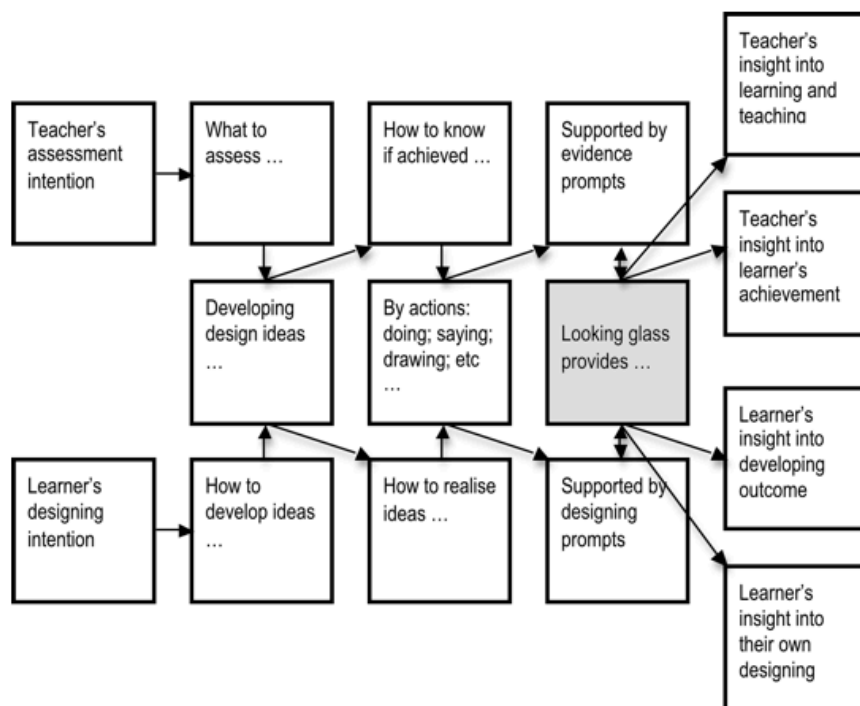

Figure 1. The mirror effect of effective evidence prompts (developed from Stables, 1992 and Kimbell & Stables, 2008b)

The growth of Technology Education across the globe has put many teachers in the position of the Early Years teachers in England and Wales when the NC was introduced. The importance of supporting linked teaching and assessment—what I am calling a pedagogic model of assessment—is critical. Moreland and Cowie (2009) support this idea further through their research on the InSiTE project, where the importance of helping teachers develop and align their PCK with their practices of Assessment for Learning were crucial in successful classroom learning and teaching. Effectively, what is being created is an authentic relationship between

learning and assessment activities. This brings us to my second guiding principle—that assessment activities should be authentic.

Maintaining authenticity activities

Authentic assessment activities are generally considered to be those where a learner is showing their understanding, skill or ability in a setting that has some validity to the nature of what is being assessed. If you want to assess how good a footballer is at scoring goals, you put them on the pitch. You don't give them a written test. The quality of 'footballing' is demonstrated through the performance or practice of the footballer. In Technology Education what is important is authenticity in technological practice and real world technology (Turnbull, 2002) or, in the words of Brown, Collins and Duguid (1989), "[t]hey need to be exposed to the use of a domain's conceptual tools in authentic activity" (p. 34)

The link to the 'real world' is important to consider in relation to the learning that can take place beyond the classroom, as we are reminded by Resnick (1987) who highlights the socially constructed, practical intelligence that is developed in the world beyond, not within, schools and classrooms. For example, Fleer and Quinones (2009), drawing on a study with primary aged children, highlight the funds of technological knowledge that learners both bring and gain through the informal setting of a school 'tinkering club' that capitalised on learners' interests in the materiality of technological artefacts and provided an experiential and social setting in which to build understanding from these interests. Both examples highlight the reality that, removed from formal classroom settings, informal authentic activities often support learning in ways that are different to formal classrooms, through practical, social, physical (hands-on) and concrete (rather than abstract) situations. The challenge is to bring the real-world authenticity of such activities into classroom-based learning and assessment. Technology Education offers real opportunities here—the 'real-worldness' of technological practice in classrooms can take learners beyond the realms of learning abstract knowledge and skills and into the social and cultural settings in which the practice takes place whilst also taking into consideration the human needs and wants that are driving the intentions behind the practice. This allows for further authenticity by engaging learners in assessment tasks that are embedded contexts that are relevant and motivating to learners, supporting their taking ownership of their learning.

Learners need plenty to get their teeth into and be challenged by. Engaging them in relevant, issues-rich tasks is a good way of allowing them to both develop and demonstrate their capability. But creating such contexts is not straightforward. There is a history of learning tasks in Technology being teacher-led and driven by the dominant culture of an education system, which may be quite different and socially and culturally irrelevant to the learners being assessed. Those devising assessment contexts can be well meaning in attempts to create motivating contexts but may be working on assumptions of what will inspire learners to show what they can do. This was highlighted in a recent study concerning the contexts in which learners in Malawi expressed an interest in the context of learning mathematics (Kazima, 2013). The study identified social relevance as

> that which connects with the present and future lives of students as well as the issues that are of importance to them and their communities, and in the interest of humankind in general. (p. 23)

and cultural relevance as defined by Ladson-Billings (1994) as education

> that empowers students intellectually, socially, emotionally and politically by using cultural referents to impart knowledge, skills and attitudes (pp. 17-18)

Kazima highlights the lack of voice of the learners, commenting that

> *Issues of relevance are often decided by a number of groups of people including policy makers, curriculum developers, textbook writers, and class teachers. Policy makers are*

often guided by the issues affecting the nation at the time. … Curriculum developers interpret the policies and develop the curriculum as they see 'relevant' to the nation's needs. Textbook writers interpret the curriculum into a possible form of teaching. Finally teachers interpret both the curriculum and the textbooks into lessons. (p. 25)

Using a survey tool previously used in an international study on relevance in maths teaching (Julie and Holtman, 2008) and customised for Malawi, Kazima goes on to show that learners who were surveyed rated contexts relating to modern technologies more highly than contexts relating to agriculture and what are termed ethnomathematics (e.g., the maths used in basket weaving). With hindsight, these findings can be explained in terms of learners' curiosity and aspirations, but given Malawi's rural context, it is clear to see how a potential mismatch can arise between the issues of relevance likely to be identified on behalf of the learners by policy makers, etc., and the very things that would inspire learners to see maths as relevant.

In piloting a parallel survey, customised to the culture of England and D&T lessons, findings had some fascinating parallels and also differences (Stables 2013, in preparation). Both groups of teenagers showed similar levels of interest in new technologies addressing issues of health and issues of secrecy and privacy. Both also showed a tendency to reject certain views of the policy makers—the Malawi maths learners rejecting agriculture and ethnomathematics, while the English D&T learners rejected learning about the lives of famous designers (popular in examination syllabi) and horticulture, a context that the current Minister for Education is keen to see adopted in D&T lessons! The lesson to be learned is that we need to focus more on how the learners themselves are engaged with selecting and building the contextual backgrounds for their tasks. This is not to deny the contribution of the teacher, but to suggest how the choice and development of authentic contexts can be optimised.

However well considered the context of an activity is, in technological practice bringing an authentic view of process is of equal importance (Moreland, 2009). When the research team in TERU was commissioned in 1985 by the UK Department for Education's Assessment of Performance Unit to assess the D&T capability of 10,000 15 year olds, one of our first challenges was to understand the design process that capability was evidenced through. Full accounts of our encounter with, and conclusion to, this challenge can be found elsewhere (Kimbell et al., 1991, Kimbell & Stables, 2008). Briefly, we came to the challenge with a distrust of the linear and cyclical models of process that were common in the literature and assessment practices at that time, because of the lack of authenticity they demonstrated. With their prescriptive and managerial structures they denied both the reality of the complex and diverse ways in which ideas are generated and the processes through which ideas grow to become thoughtful and well-developed working realities. Drawing on our experience as designers and teachers, alongside the performance data collected through our research, we created a different model where the line of development was from hazy to clear and the process was driven by the iterative interaction of action and reflection (see Figure 2). This model allowed us to understand the various ways in which the 10,000 learners in question went about their processes of designing and (along with a team of 100 teacher assessors) come to conclusions about the learners' levels of performance in D&T.

THE INTERACTION OF MIND AND HAND

IMAGING AND MODELLING INSIDE THE HEAD — CONFRONTING REALITY OUTSIDE THE HEAD

HAZY IMPRESSIONS — DISCUSSION, DRAWINGS, SKETCHES, DIAGRAMS, NOTES, GRAPHS, NUMBERS

SPECULATING AND EXPLORING — MODELLING IN SOLID TO PREDICT OR REPRESENT REALITY

CLARIFYING AND VALIDATING — PROTOTYPING OR PROVISIONAL SOLUTIONS

CRITICAL APPRAISAL

THE POTENTIAL OF MORE DEVELOPED THINKING — THE POTENTIAL OF MORE DEVELOPED SOLUTIONS
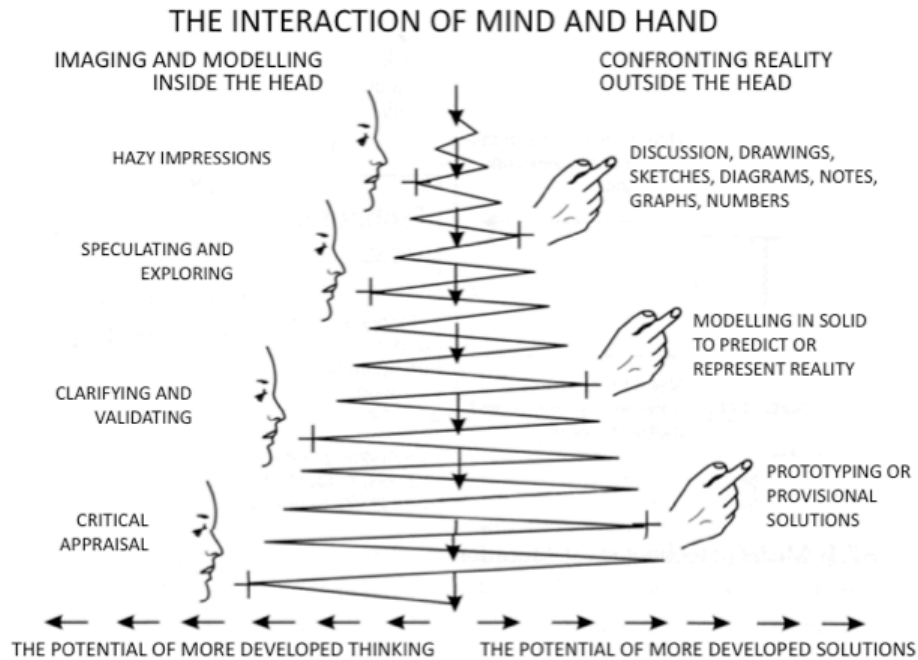
Fig 1 The APU Design and Technology Model (Kimbell et al., 1991, p. 20)

While over the years we have come to have richer understandings of the ways individuals approach processes of designing, not least through others exploring similar territory (e.g. Buchanan, 1995; Darke, 1979; Lawson, 1990, 2004; Middleton, 2005; Nelson & Stolterman, 2003) we have continued to find the model useful in understanding the reality of the complex processes at work. This same model of process continues to underpin our work on assessment, including our current work on digital assessment portfolios. Whilst the model is challenging from a manageability point of view, assessment that respects the authenticity of individual processes is important in considering authenticity from a learner's perspective. The route one learner takes through the process will be different to another and a teacher accepting the iterative and responsive nature of the model needs alternative strategies to manage the assessment and, by implication, the learning. Our experience from research suggests that the pivot points in the activity are action and reflection: supporting learners to explore their thinking through action and then pausing to reflect on that action is a fundamental rhythm to encourage.

The final piece in this jigsaw of authenticity is of the criteria against which assessment judgements are made. This takes us back to the example given earlier of assessing a footballer's ability to score goals. The criteria need to be 'authentic' in terms of what they are attempting to reveal and they need to be applied in a situation where they can be validly evidenced. "Can score goals" is unlikely to be evidenced fully or validly in a written test.

Reflecting back to the time when we created the APU D&T model of process, our hunch was that criteria would be derived from key elements of the process (identifying and addressing issues in the task; having a grip on generating ideas and developing solutions; appraising their thinking with a sound, critical eye) and interconnectedness of the iteration between thought and action. Moreover, with a belief that the value of a whole enterprise is greater than the sum of its parts, our 'hunch' was to view all evidence holistically. Taking account of the whole of what a learner has set out to do, how they have gone about it, and what they have achieved, enabled us to see elements of the process (evaluating, generating, researching, problem solving, etc.) at whatever stage they appeared and to take an overall position on the learner's achievements. Looking within the evidence allowed us to diagnose other aspects—strengths and weaknesses of their

approach, how they had understand the application of knowledge, etc. This approach may seem counter-intuitive but, for the record, the most statistically reliable judgements the teachers (assessors) made were the holistic ones. The smaller and more atomised the assessment decisions became, the less statistically reliable they were found to be (Kimbell et al., 1991).

The importance of judgement

Our research has consistently indicated that engaging in holistic assessment acts as an important professional development tool for teachers and an important learning tool for learners. A facet of this is the emphasis on making judgements of qualities, as opposed to awarding marks based on 'right' or 'wrong' answers. Boud (2007b) suggests that judgement is "the capacity to evaluate evidence, appraise situations and circumstances astutely, to draw sound conclusions and act in accordance with this analysis" (p. 19).

Making a judgement is a complex process, but that does not necessarily mean that it is difficult or unreliable. Judgements are not made in a vacuum and, in educational contexts, often involve drawing on an individual's repertoire of professional experience. It has been our experience that when a teacher or learner is asked to make an overarching judgement about what has been achieved in an assessment task this allows them to reflect carefully on the evidence and the context in which it has been created. If asked to share and debate their judgements, whether in assessment moderation meetings or in self and peer assessment activities, their thinking becomes clearer as they articulate and justify their position. It is often in this dialogue that teachers and learners grow in both their understanding and confidence.

Boud (2007b) identifies the act of making a judgment as moving an individual from passive recipient to active assessor. Where learners are concerned he sees a double value—learning to form judgements as well as learning to act on the judgements formed. Drawing on both social theory (Bourdieu & Wacquant, 1992; Giddens, 1991) and psychology (Karoly, 1993) he goes further to propose that bringing judgement to the fore in assessment helps develop both reflexivity and self-regulation in learners as they enable individuals

> to 'look again', to monitor one's own performance, to see one's own learning in the context in which it is deployed and to respond with awareness to the exigencies of the tasks in which one is engaged. Reflexivity and self-regulation … involve dispositions and an orientation to both work and learning. They also have an affective dimension. They involve confidence and an image of oneself as an active learner, not one solely directed by others. A focus on reflexivity and self-regulation is a key element in constructing active learners, as these features need to be constructed by both teachers and examiners themselves. (pp. 21-22)

The approach to using judgement that we have taken in recent research (the e-scape project) has been based the law of comparative judgment (Thurstone, 1927), which, at its simplest level, is based on detailed, holistic reviewing of two different pieces of work and then making a judgement about which is better. Linked to a computer-driven algorithm that presents 'judges' with series of pairs of work to be judged, a rank is created. Our development and use of this process is described in detail elsewhere (e.g., Kimbell et al., 2009). As an assessment process, statistically it has very, very high reliability. But for our purposes here, it also presents potential for professional development of teachers and metacognitive understanding for learners, as a concrete example of the reflexivity and self-regulation identified by Boud. In a pilot project, fifteen year olds who had taken part in the e-scape assessment activity then became part of the judging team. Not only were their judgements consistent with the adults, they found the exercise highly illuminating because of the insights gained from assessing each other's work. They commented that they felt better prepared for future work (Kimbell, 2012). In summing up their experience, their teacher (who had also engaged in the research as a judge) commented

> It was unbelievable how quickly all of the students managed to get to grips with the assessment process. And listening in to the conversations that were going on during the

judgments it was apparent that they didn't have many of the hang-ups that I as a teacher have experienced. Without a doubt they were able to spot the creative and innovative thinking in the design work and were rarely taken in by "pretty" or "content free" products. (Kimbell et al., 2009, p. 161)

Not only does this example illustrate the value of engaging learners in holistic judging as a form of peer and self assessment, it also indicates new possibilities in democratic approach in the context of high stakes assessment, this pilot being part of a trial for authentic assessment that could be used in high stakes.

The importance of equity

As an example of democratic assessment, the previous example places the assessed in the active role of the assessor. But not all assessment systems are so inclusive. To be equitable, assessment practices need to ensure the assessment process itself is fair, for example, taking account of the ways different learners might demonstrate achievement and attainment, including through valid but less tangible modes, such as group work and talk. Referring back to Shepard's (2000) concern that assessment practices have not progressed in step with understanding of effective ways of learning, particularly through sociocultural approaches, there is a real danger that assessment is inherently unfair, and therefore invalid, if misaligned with the real evidence of learning.

Ensuring that learners have the opportunity to show their achievements in appropriate ways requires shifts in thinking around the nature of evidence. For learners with special educational needs it may be something of a worthless task to attempt to assess their understanding through a written test if this is not the best way for them to communicate. For learners whose first language is not that of the dominant educational culture, then assessing them through this language is unlikely to provide a true reflection of what they can achieve. Both of these examples come from what might be seen as 'special' situations. But given the nature of Technology Education and the ways in which technological literacy and capability can be enacted, and given the understandings we now have about preferred learning styles and even designing styles (Lawler, 1999, 2006) it seems not just iniquitous but also inefficient not to take these into account when considering effective assessment approaches. With the increased possibilities of digital tools for capturing data in diverse ways there is a real opportunity to use these tools to support learners to communicate the evidence of their technological practice in ways that genuinely and appropriately demonstrate their capability.

Impact of digital technologies on assessment

Digital tools can be used effectively to support assessment of different learning and designing styles. But just because a tool is digital, their use in assessment doesn't, by definition, make the approach better. If we use digital tools to do things that were wrong in an analogue world, they won't become right just because of the use of 'new' technologies. However when considering the affordances of digital technologies, including in relation to the underpinning principles that have been outlined above, it can be seen that they offer positive benefits to various aspects of assessment processes and practices.

An immediate benefit is the way evidence can be organised in digital structures, an approach used quite extensively in e-portfolios. Presenting both the process and the outcome of technological activity digitally allows a rich collection of text, image, audio and video to be included. E-portfolios are increasingly commonplace, even for high-stakes assessment. Pragmatically they allow for the submission, storage and archiving of assessment evidence in an accessible way that isn't physical space hungry. Creating a summative assessment portfolio can be a useful activity in itself as the learner reviews and selects work to submit. But this approach can fall into the same trap as when it is done through pencil and paper if valuable learning time is wasted in re-presenting work for the assessor, denying the real benefits of dynamic digital capture using digital tools to 'hoover up' the evidence as it is generated. This latter approach

draws on the range of tools that are available for documenting written and spoken word alongside image-based data.

The mobility of digital tools, with smart phones, netbooks and laptops, also can contribute to effective and authentic approaches to assessments since the digital evidence collection can be undertaken in a range of settings, rather than being tied to regular classroom activity or examination halls. For example, assessments that are in the workplace setting can readily collect evidence of the performance of activities in the workplace. This benefit can also be seen for activities that are best captured through their physical enactment where learners are demonstrating their understanding 'for real' either through practical activities or the presentation of their ideas.

As we have seen earlier, there are benefits to using the practice of holistic judgement in assessing project work. When linked to web-based systems, digital tools can make this process more manageable, allowing multiple assessments to take place at any given time and in multiple locations (Kimbell et al., 2009). Web-based digital portfolios can also be hugely beneficial for formative assessment, as we found when trialing the e-scape system in a number of schools in Israel (Stables & Lawler, 2011). Learners felt that their teachers had a better understanding of the process they had gone through in their projects and teachers felt that it benefitted a broader range of learners, even the special needs children finding assessment more supportive. Seery, Canty and Phelan (2012) have also shown how the approach proved both valid and reliable in undergraduate peer assessment of design project work.

Each of the above dimensions of using digital tools in assessment has the added value of supporting flexible and agile learning. In the context of the dynamic nature of technological projects, digital tools have the dual advantage of supporting the learning itself whilst also documenting the evidence as it is created. Digital tools can play an important role in supporting future approaches to assessment in Technology Education by adding value to existing effective approaches and by the new functionalities that they bring to help address age-old problems such as reliability of judgements in performance settings.

The future agenda

Through this chapter a range of issues, ideas and insights have been presented. History has shown the damage done in the name of assessment but research has provided some clarity and grounding that provides direction to positive avenues for moving forward.

It has been particularly useful to be reminded of how the paradigm that dominates assessment practices still largely conflicts with more progressive understandings in learning and teaching. This gives a clear message about the importance of leading assessment developments through the lenses of more progressive paradigms of learning and teaching and helping all stakeholders, and especially teachers, to understand both the theoretical and practical underpinnings to move Technology Education assessment to more constructivist and sociocultural approaches. It is interesting to note how, in Technology Education research, developments in taking socio-cultural approaches to assessment are largely in the field of early years and primary education. Much has to be gained from extending this research to secondary phases, which have been more dominated by high-stakes summative assessment. Placing a greater emphasis on social and cultural relevance in assessment activities will also open up new understandings to support more equitable approaches.

The chapter has also indicated the potential affordances of new technologies in assessment and this is any area that will inevitably play an increasing role. An important message here is to ensure that digital approaches build on and develop existing sound and authentic approaches such that these are enhanced. This links to the overarching message of the chapter—to continue to push for assessment that is authentic on all levels.   The practice-based, real-world nature of Technology Education provides an excellent, sympathetic setting for the development of authentic approaches to assessment. Core to future developments should be continued support

for teachers to align learning and assessment in order to engage and motivate learners through technological practice that makes visible to the learners and their teachers and assessors the learning that has taken place and the capability that has been developed.

References

Anning, A., Cullen, J., & Fleer, M. (2009). Early childhood education: society and culture (2nd Edition ed., pp. 238). London, Thousand Oaks, New Delhi, Singapore: Sage.

Atkinson, S. (2000). Does the need for high levels of performance curtail the development of creativity indesign and technology project work? *International Journal of Technology and Design Education, 10*(3), 255-281.

Beets, P., & van Louw, T. (2011). Social justice implications of South African school assessment practices. *Africa Education Review, 8*(2), 16.

Biggs, J. (1996). Enhancing teaching through constructive alignment. *Higher Education, 32*, 347-364.

Boud, D. (2000). Sustainable assessment: Rethinking assessment for the learning society. *Studies in Continuing Education, 22*(2), 26.

Boud, D. (2007). Reframing assessment as if learning were important. In D. Boud & N. Falchikov (Eds.), *Rethinking Assessment in Higher Education: Learning for the longer term* (pp. 14-26). Abingdon: Routledge.

Boud, D., & Falchikov, N. (2007). Introduction: assessment for the longer term. In D. Boud & N. Falchikov (Eds.), *Rethinking Assessment in Higher Education: Learning for the longer term* (pp. 3-13). Abingdon: Routledge.

Bourdieu, P., & Wacquant, L. (1992). *An invitation to reflexive sociology*. Chicago: University of Chicago.

Brown, J. S., Collins, A., & Duguid, P. (1989). Situated cognition and the culture of learning. *Educational Researcher, 18*(1), 32-42.

Buchanan, R. (1995). Wicked problems in design thinking. In V. Margolin & R. Buchanan (Eds.), *The idea of design* (pp. 3-20). Cambridge, MA; London, England: MIT.

Darke, J. (1979). The primary generator and the design process. *Design Studies, 1*(1), 36-44.

DES/WO. (1988). *National Curriculum: Task group on assessment and testing: A report*. London: Department of Education and Science and the Welsh Office.

Elton, L. (2006). Assessing creativity in an unhelpful climate. *Art, Design & Communication in Higher Education, 5*(2), 119-130.

Fleer, M., & Quinones, G. (2009). Assessment of children's technological funds of knowledge as embedded community pracctices. In A. Jones & M. de Vries (Eds.), *International handbook of research and development in Technology Education* (1st ed., pp. 477-492). Rotterdam: Sense.

Fleming, R. (2013). *Design Education for a sustainable future*. London, New York: Earthscan (Routledge).

Giddens, A. (1991). *Modernity and self-identity: Self and society in the late modern age*. Cambridge: Polity.

Harlen, W., & Deakin Crick, R. (2002). A systematic review of the impact of summative assessment and tests on students' motivation for learning (EPPI-Centre Review, version 1.1) *Research Evidence in Education Library.* (Vol. Issue 1). London: EPPI-Centre, Social Science Research Unit, Institute of Education.

Harris, M., & Wilson, V. (2003). Designs on the curriculum? A review of the literature on the impact of Design and Technology in schools in England. Nottingham: Department for Education and Skills

Julie, C., & Holtman, L. (2008). The Relevance of School Mathematics Education (ROSME). In L. Holtman, C. Julie, O. Mikalsen, D. Mtetwa & M. Ogunniyi (Eds.), *Some developments in*

research in science and mathematics in Sub-Saharan Africa: Access, relevance, learning and curriculum (pp. 379-405). Somerset West, SA: African Minds.

Karoly. (1993). Mechanisms of self-regulation: A systems view. *Annual Review of Psychology, 44*, 23-52.

Kazima, M. (2013). *Relevance and school mathematics*. Paper presented at the SAARMSTE 2013: Making Mathematics, Science and Technology Education, Socially and Culturally relevant in Africa, University of the Western Cape, Cape Town.

Kelly, A. V. (2009). *The curriculum: Theory and practice* (6th Edition ed.). Los Angeles, London, New Delhi, Singapore, Washington DC: Sage.

Kimbell. (2012). Evolving project e-scape for national assessment. *International Journal of Technology and Design Education, 22*(2), 135-155.

Kimbell, R. (1997). *Assessing technology: International trends in curriculum and assessment*, Buckingham, UK: Open University Press

Kimbell, R. (2006). Innovative performance and virtual portfolios - a tale of two projects. *Design and Technology Education: An international journal, 11*(1), 18-30.

Kimbell, R., & Stables, K. (2008a). Designerly thinking beyond design. In V. Lindberg & K. Borg (Eds.), *Kunskapande, kommunikation och bedömning: i gestaltande utbildning* (pp. 188-198). Stockholm: Stockholms universitets förlag.

Kimbell, R., & Stables, K. (2008b). *Researching design learning: Issues and findings from two decades of research and development*. Berlin: Springer.

Kimbell, R., Stables, K., Wheeler, T., Wozniak, A., & Kelly, A. V. (1991). *The assessment of performance in design and technology*. London: SEAC / HMSO.

Kimbell, R., Wheeler, T., Stables, K., Shepard, T., Martin, F., Davies, D., . . . Whitehouse, G. (2009). e-scape portfolio assessment: a research & development project for the Department of Children, Families and Schools, phase 3 report (pp.169). London: Goldsmiths, University of London.

Ladson-Billings, G. (1994). *The dreamkeepers: Successful teachers for African-American children*. San Francisco: Jossey-Bass.

Lawler, T. (1999). Exposing the gender effects of design and technology project work by comparing strategies for presenting and managing pupils' work. In P. H. Roberts & E. W. L. Norman (Eds.), *IDATER 99: International Conference on Design and Technology Educational Research and Curriculum Development*. (pp130-137) Loughborough: Loughborough University of Technology.

Lawler, T. (2006). *Design styles and teaching styles: A longitudinal study of pupils' ways of doing designing following complementary re-grouping and teaching.* Paper presented at the TERC 2006: Values in Technology Education, Gold Coast, Australia.

Lawson, B. (1990). *How designers think: The design process demystified* (2nd ed.). Oxford: Butterworth Architecture.

Lawson, B. (2004). *What designers know*. Oxford: Elsevier.

MacLeod, R. M. (1982). *Days of judgement: Science, examinations and the organization of knowledge in late Victorian England*. Driffield: Nafferton Books.

Middleton, H. (2005). Creative thinking, values and design in Technology Education. *International Journal of Technology and Design Education, 15*(1), 61-71.

Moreland, J. (2009). Assessment: focusing on the learner and the subject. In A. Jones & M. de Vries (Eds.), *International Handbook of research and development in Technology Education* (1st ed., pp. 445-448). Rotterdam: Sense.

Moreland, J., & Cowie, B. (2009). Making meaning in primary technology classrooms through assessment for learning. In A. Jones & M. de Vries (Eds.), *International handbook of research and development in Technology Education* (1st ed., pp. 461-476). Rotterdam: Sense.

NCEE. (2007). *Tough Choices or Tough Times: the report of the new Commission on the Skills of the American Workforce* (Vol. Executive Summary). Washington: National Center on Education and the Economy.

Nelson, H. G., & Stolterman, E. (2003). *The design way: Intentional change in an unpredictable world*. Englewood Cliffs, NJ: Educational Technology Publications.

Ofqual. (2013). *Review of Controlled Assessment in GCSEs*. London: Office of Qualifications and Examinations Regulations

Ofsted. (2001). *Ofsted subject reports 1999 - 00: Secondary Design and Technology*. London: Department for Education and Employment.

Ofsted. (2002). *Secondary subject reports 2000/01: Design and technology*. London: Department for Education and Employment.

Pellegrino, J. W. (2006). *Rethinking and redesigning curriculum, instruction and assessment: what contemporary research and theory suggests. A paper commissioned by the National Center on Education and the Economy for the New Commission on the Skills of the American Workforce*. Washington: National Center on Education and the Economy.

Prest, D. (2002). *An analysis of the attainment target level descriptors and associated programme of study in relation to the design and technology mission statement*. London: Department for Education and Skills, Design and Technology Strategy Group.

Resnick, L. B. (1987). Learning in school and out. *Educational Researcher, 16*(5), 8.

Seery, N., Canty, D., & Phelan, P. (2012). The validity and value of peer assessment using adaptive comparative judgement in design driven practical education. *International Journal of Technology and Design Education, 22*(2), 21.

Shepard, L. A. (2000). The role of assessment in a learning culture. *Educational Researcher, 29*(7), 4-14.

SSEC. (1960). *The Beloe Report: Secondary schools examinations other than the GCE*. Lodnon: HMSO.

Stables, K. (1992). The assessment of technology at key stage 1. In C. Gipps (Ed.), *Developing assessment for the National Curriculum* (pp. 42 - 50). London: Kogan Page.

Stables, K. (2013, December). *Social and cultural relevance in approaches to developing designerly well-being: the potential and challenges when learners call the shots in Design and Technology projects*. Paper presented at the Technology Education for the future: A play on sustainability, Christchurch, New Zealand.

Stables, K., & Lawler, T. (2011). *Assessment in my palm: e-scape in Israel Evaluation of Phase 1* (pp. 18). London: TERU: Goldsmiths, University of London.

Taunton. (1868). The Taunton Report. London: Schools Inquiry Commission

Thurstone, L. L. (1927). A law of comparative judgement. *Psychological Review, 34*, 273-286.

Turnbull, W. (2002). The place of authenticity in Technology in the New Zealand curriculum. *International Handbook of Technology and Design Education, 12*, 23-40.