

DESIGNING FORMATIVE ASSESSMENT LESSONS FOR CONCEPT DEVELOPMENT AND PROBLEM SOLVING

Malcolm Swan
University of Nottingham
malcolm.swan@nottingham.ac.uk

Formative assessment is the process by which teachers and students gather evidence of learning and then use this to adapt the way they teach and learn. I describe a design research project in which we integrated formative assessment strategies into lesson materials that focus on developing students' conceptual understanding and their capacity to tackle non-routine problems. A theoretical framework for assessment task design is presented, together with an analysis of research-based principles for formative assessment lesson design. Particular aspects are highlighted: the roles of pre-assessment, formative feedback questions and sample work for students to critique. While there are some early signs that these lessons provide an effective model for teachers to introduce formative assessment into everyday classroom practice, the materials require a radical shift in the predominant culture within most classrooms.

Keywords: Assessment and Evaluation; Design Experiments

Introduction

There is little doubt that assessment has a profound impact on the nature of student learning, and that this is often detrimental in nature. Our assessment practices have the potential to convey our valued learning goals to students, but this is often unrealized because the tasks and methods we use do not reflect these values. It has been found, for example, that even when teachers clearly acknowledge the importance of eliciting students' understanding and of giving useful, qualitative feedback, the tests they use encourage 'rote and superficial learning' and appear more concerned with grading and record keeping than with developing learning (Black & Wiliam, 1998). The poor design of summative, high-stakes tests must take some of the blame for this. These are designed to be cheap, predictable and simple to grade and, in consequence, focus on fragments of mathematical performance. Policy makers tend to ignore their powerful backwash effect and continue to claim that tests are merely measuring instruments (ISDDE, 2012).

Assessment needn't be this way. High quality assessment, focused on important mathematics, can be a powerful lever for positive change. This requires a radical shift away from multiple choice, computer-based assessments of procedural knowledge toward assessments that focus on the mathematics we care about - understanding, reasoning and problem solving. More substantial assessment tasks are required and scoring must begin to assess the quality of students' extended reasoning. (This is possible even in high stakes assessment when human judgment, rather than machine scoring, is allowed to have a role. Point scoring rubrics of chains of reasoning, long established in other subjects, can give reliable scores on mathematics tests. Reliable qualitative methods, such as adaptive comparative judgment, are also now recognized as a possible way forward (Jones, Pollitt, & Swan, 2015). Further, when teachers are involved in scoring, suitably organized, it can have considerable value for professional development.)

In this paper, however, I have insufficient space for a thorough discussion of high stakes assessment. Instead I wish to focus on the potential of classroom assessment to produce significant and substantial student learning gains. This potential was brought to our attention by the research reviews of Black, Wiliam and others (Black, Harrison, Lee, Marshall, & Wiliam, 2003; Black & Wiliam, 1998; Black & Wiliam, 1999). In their original definition, the term 'formative assessment' is taken to include:

Bartell, T. G., Bieda, K. N., Putnam, R. T., Bradfield, K., & Dominguez, H. (Eds.). (2015). *Proceedings of the 37th annual meeting of the North American Chapter of the International Group for the Psychology of Mathematics Education*. East Lansing, MI: Michigan State University.

... all those activities undertaken by teachers, and by their students in assessing themselves, which provide information to be used as feedback to modify the teaching and learning activities in which they are engaged. Such assessment becomes ‘formative assessment’ when the evidence is actually used to adapt the teaching work to meet the needs. (Black & Wiliam, 1998, p. 140)

This definition is wide-ranging, and includes both pre-planned and incidental assessment activities, such as diagnostic tests, oral questioning, collaborative tasks and observation of students. Improving the nature and focus of teacher-student and student-student communication is central. Most importantly, however, it must lead to adaptive action, not just the reteaching of the material concerned.

Since their work was published, this definition has often been mutated to mean more frequent testing, scoring and record keeping. In the UK, for example, one government initiative, “Assessing Pupil Progress” (APP) degenerated into the atomized profiling of pupils. This involved teachers in monitoring work, keeping files on pupils and regularly assessing progress against detailed criteria. Teacher workload was significantly increased and many teachers did not use the feedback to improve instruction. Recognizing such mutations, Black and Wiliam refined their definition a little differently in a later paper, laying more emphasis on the *agents* in the process: teachers, learners and peers, and the requirement for each of these agents to make *effective use* of the evidence obtained:

Practice in a classroom is formative to the extent that evidence about student achievement is elicited, interpreted, and used by teachers, learners, or their peers, to make decisions about the next steps in instruction that are likely to be better, or better founded, than the decisions they would have taken in the absence of the evidence that was elicited. (Black & Wiliam, 2009, p. 9)

The interaction between these agents and the three main aspects of formative assessment: identifying where learners are in their learning, where they are going, and how to bridge the gap have been clearly articulated by Wiliam, and Thompson (2007), see Table 1. Within the matrix formed, are their five “key strategies” of formative assessment.

Table 1: Key Strategies of Formative Assessment

	Where the learner is going	Where the learner is right now	How to get there
Teacher	1. Clarifying learning intentions and criteria for success	2. Engineering effective classroom discussions and other learning tasks that elicit evidence of student understanding	3. Providing feedback that moves learners forward
Peer	Understanding and sharing learning intentions and criteria for success	4. Activating students as instructional resources for one another	
Learner	Understanding and sharing learning intentions and criteria for success	5. Activating students as the owners of their own learning	

Black and Wiliam launched programs of work that aimed at engaging teachers in these key strategies, but found that regular meetings over a period of years were needed to enable a substantial proportion of teachers to acquire the “adaptive expertise” (Hatano & Inagaki, 1986) needed for self-directed formative assessment. This is clearly an approach that is challenging to implement on a large scale.

The Mathematics Assessment Project

In 2009, the Bill & Melinda Gates Foundation approached us at Nottingham to develop a suite of “formative assessment lessons” to form a key element in the Foundation’s program for “College and Career Ready Mathematics” based on the Common Core State Standards for Mathematics (NGA & CCSSO, 2010). In response, the Mathematics Assessment Project (MAP) was designed to explore

Bartell, T. G., Bieda, K. N., Putnam, R. T., Bradfield, K., & Dominguez, H. (Eds.). (2015). *Proceedings of the 37th annual meeting of the North American Chapter of the International Group for the Psychology of Mathematics Education*. East Lansing, MI: Michigan State University.

how far well-designed teaching materials can enable teachers to make high-quality formative assessment an integral part of the implemented curriculum in their classrooms, even where linked professional development support is limited or non-existent. The lessons are thus designed, not only to provide teachers with diagnostic information, but to enable them use it to move each student's reasoning forward.

To date, we have designed and developed about a hundred formative assessment lessons to support US Middle and High Schools in implementing the new Common Core State Standards for Mathematics (CCSSM). Each lesson consists of student resources and an extensive teacher guide. The data we have does appear to support the assertion that these lessons have enabled teachers to integrate the key strategies for formative assessment, as identified in Table 1, into their normal teaching. The research-based design of these lessons, now called Classroom Challenges, forms the focus of this paper.

A Design-Based Methodology

Our methodology for lesson design was based on design research principles, involving theory-driven iterative cycles of design, enactment, analysis and redesign (Barab & Squire, 2004; Bereiter, 2002; Cobb, Confrey, diSessa, Lehrer, & Schauble, 2003; DBRC, 2003; Kelly, 2003; van den Akker, Graveemeijer, McKenney, & Nieveen, 2006). In contrast with much design research, we worked to ensure that the products were robust in large-scale use by fairly typical end-users. This is, in fact why some prefer the term "engineering research" to design research (Burkhardt, 2006). Each lesson was developed, through two iterative design cycles, with each lesson trialed in three or four US classrooms between each revision. This sample size enabled us to obtain rich, detailed feedback, while also allowing us to distinguish general implementation issues from more idiosyncratic variations by individual teachers. As we were designing at a distance, revisions had to be based on structured, detailed feedback from experienced local observers in California, Rhode Island and the Midwest. We obtained approximately 700 observer reports of lessons, from over 100 teachers (over 50 schools) using these materials. We also observed many of the lessons first-hand, in UK schools.

In order for feedback to be useful in the revision process it had to be specific and reliable, based on a *detailed description* of what happened in each lesson. To meet this challenge, a protocol was developed. Two design questions permeated the protocol: How well did the materials communicate the formative assessment strategies to the teacher? How far was the learning experience profitable for students? The protocol was in three parts. The first part was descriptive, asking for the context, the nature of the students, the environment, the support given to the teacher, followed by a vivid description of the course of the lesson, illustrated by a sample of student work of varied quality. Significant events that might inform the designer were noted. The second part was analytical. Observers were asked for: their overall impressions; deviations from the lesson plan; quality of teacher questioning; quality of student reasoning, explanations, discussion and written work. They were also asked to provide evidence of learning. They were specifically asked about the relevance of the formative assessment opportunities. The third part sought the teacher's views, through an interview after the lesson. Teachers were asked about their lesson preparation, their views on the lesson plan, the lesson and the response of students, and implications for professional development. In developing 100 Classroom Challenges over the course of the project, about 700 such reports were obtained and discussed by the design team. This process enabled us to obtain rich, detailed feedback, while also allowing us to distinguish general implementation issues from idiosyncratic variations by individual teachers. On this basis the lessons themselves were revised, and ultimately published on the web: <http://map.mathshell.org.uk/materials/index.php>.

Theoretical Framework for Assessment Task Design

Our first priority was to clarify the learning intentions for *Classroom Challenges*. The CCSSM make it clear that the goals of the new curriculum are to foster a deeper, connected conceptual understanding of mathematics, along with the strategic skills necessary to tackle non-routine problems. A particular emphasis is the development of mathematical *practices* that should permeate all mathematical activity. We rapidly found it necessary to distinguish between tasks that are designed to foster conceptual development from those that are designed to develop problem-solving strategies. In the former, the focus of student activity is on the analysis and discussion of different *interpretations* of mathematical ideas, while in the latter the focus is on discussing and comparing alternative *approaches* to problems.

The intention was that concept lessons might be used partway through the teaching of a particular topic, providing the teacher with opportunities to assess students' understanding and time to respond adaptively. Problem solving lessons were designed to be used more flexibly, for example between topics, to assess how well students could select already familiar mathematical techniques to tackle unfamiliar, non-routine problems and thus provide a means for improving their strategic thinking.

The validity of any assessment scheme lies in the design of the tasks, which should reflect the intentions of the curriculum in a balanced way. We therefore begin by describing our task design framework. This is followed by a review of the research we used to design the formative assessment lesson structures within which the tasks are embedded.

(i) Assessment Task Genres for Concept Development

The tasks we selected for concept *Classroom Challenges* were designed to foster collaborative sense-making. Sierpiska (1994) suggests that people feel they have understood something when they have achieved a sense of order and harmony, where there is a sense of a 'unifying thought', of simplification, of seeing an underlying structure and that in some sense, feeling that the essence of an idea has been captured. She lists four mental operations involved in understanding: "identification: we can bring the concept to the foreground of attention, name and describe it; discrimination: we can see similarities and differences between this concept and others; generalisation: we can see general properties of the concept in particular cases of it; synthesis: we can perceive a unifying principle." To this, we would add the notions of representation. When we understand something, we are able to represent it in a variety of ways: verbally, visually, and/or symbolically. In the light of this, we developed four 'genres' of tasks for our concept development lessons (Table 2).

Space dictates that we only provide a few examples. For *Classify and define*, students were typically invited to sort a collection of cards showing mathematical objects using their own, or given criteria. The results of their sorting were then offered to other students, who would reconstruct the criteria that had been used. The objects ranged from geometric shapes to algebraic functions. As Zaslavsky (2008) has shown, this is a powerful way of enumerating properties of mathematical objects. Occasionally, students were presented with a mathematical object and were invited to list as many of its properties as possible. The task then became: "do any of these properties, taken individually, *define* the object?" or "do any *pairs* of these properties define the object?" (Figure 1). This resulted in a search for justifications and counterexamples. (This could be very demanding. For example, consider the pair of statements: "When $x = 0$, $y = 0$ "; "When x doubles in value, y doubles in value". Do these statements *define* proportion? If not, then find a function that satisfies these statements but is not a proportion). Seeking definitions in this way lies at the very heart of mathematical activity (Lakatos, 1976).

Table 2: Assessment Task Genres for Concept Development

Assessment task genres	Sample classroom activities.
Classify and define mathematical objects and structures.	Identifying and describing attributes and sorting objects accordingly. Creating and identifying examples and non-examples. Creating and testing definitions.
Represent and translate between mathematical concepts and their representations.	Interpreting a range of representations including diagrams, graphs, and formulae. Translating between representations and studying the co-variation between representations.
Justify and/or prove mathematical conjectures, procedures and connections.	Making and testing mathematical conjectures and procedures. Identifying examples that support or refute a conjecture. Creating arguments that explain why conjectures and procedures may or may not be valid.
Identify and analyze structure within situations	Studying and modifying mathematical situations. Exploring relationships between variables. Comparing and relating mathematical structures.

<i>Mathematical object</i>	<i>A square</i>	<i>A proportional relationship exists between two continuous variables x and y.</i>
<i>Properties</i>	Four equal sides Two equal diagonals Four right angles Two pairs of parallel sides Four lines of symmetry	The graph of y against x is linear. $y \div x$ always gives the same result. When $x = 0$, $y = 0$ When x doubles in value, y doubles in value When x increases by equal steps then so does y

Figure 1: Observe, Classify and define: Listing properties and building definitions

For *represent and translate*, we developed activities that require students to translate between numerical, verbal, graphical, algebraic and other representations. Typically, groups of students were given collections of cards that they were asked to sort according to whether or not the cards convey equivalent representations. Common misinterpretations were foregrounded by including translations that are commonly confused. For example, students were given a collection of four money cards (\$100; \$150; \$160; \$200) and a collection of ten ‘arrow’ cards showing percentage increase and decrease (e.g. “up by 25%”; “down by 25%”). They were asked to place the money cards in a square formation and place the percentage cards between them in appropriate places (Figure 2 shows just one side of the ‘square’). Typically, students considered “up by 25%” and “down by 25%” to be inverse statements and placed them together between the money cards \$160 and \$200. Subsequently, the teacher introduced further arrow cards showing “decimal multipliers” (e.g. $\times 1.25$; $\times 0.8$). As students place these, they checked both with a calculator and by relating them to the percentage cards

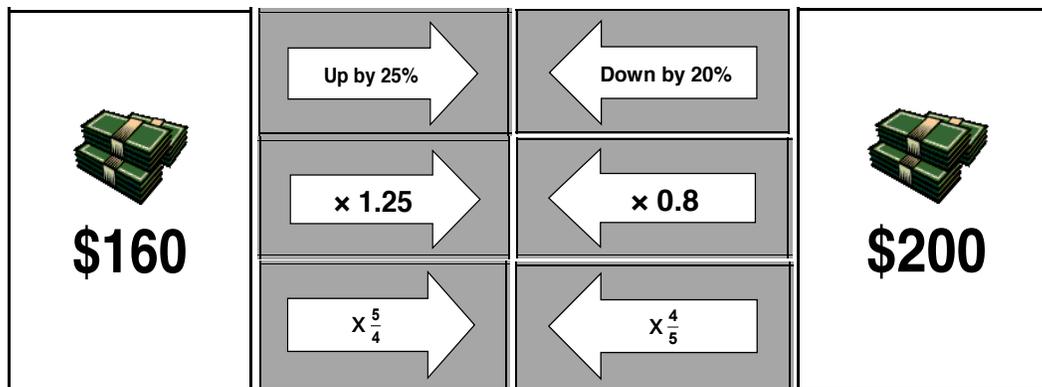


Figure 2: Represent and translate: Percentage increase and decrease.

already in position. This created “cognitive conflict” and discussion as inconsistencies were found. Later, further cards were added, as shown. Connections were drawn between all these representations and generalizations were made.

For *justify or prove* category, we designed collections of conjectures, and it was the students’ task to determine their domains of validity. Figure 3 illustrates a typical selection of such assertions.

<p>Pay rise Max gets a pay rise of 30%. Jim gets a pay rise of 25%. So Max gets the bigger pay rise.</p>	<p>Fractions If you add the same number to the numerator and denominator of a fraction, the fraction will increase in value.</p>
<p>Area and perimeter When you cut a piece off a shape you reduce its area and perimeter.</p>	<p>Right angles A pentagon has fewer right angles than a rectangle.</p>
<p>Diagonals The diagonals of a quadrilateral divide the quadrilateral into 4 equal areas.</p>	<p>Right triangle If a right-angled triangle has integer sides, the incircle has integer radius.</p>

Figure 3: Justify or prove: A selection of conjectures to test.

Normally, a set of cards was related to a single mathematical topic, and contained some commonly held beliefs. Students were instructed: “If you consider a statement to be always true or never true, then try to explain clearly how you be sure. If you think a statement is sometimes true, then try to describe all the cases when it is true and all the cases when it is false.” Thus students had first to identify the variables involved and then test the assertion by constructing examples and counterexamples. In some cases a formal proof could be sought. When students became stuck, the teacher pointed them toward particular cases to test. For example, in *Diagonals*, students often claimed that the statement is true for squares, but not for rectangles. The teacher needed to prompt them to re-consider and then go on to study a wider range of quadrilaterals to try to find *all* cases where the statement was valid.

Finally, we turn to *identify and analyze structure*. When students had tackled a conventional word problem, for example, they were invited to analyze its structure and in so doing construct further problems. The problem was rewritten as a list of variables together with their original values, including the solution to the original problem (see Figure 4). The task was to first describe how each variable might be obtained from the others, then to explore the effect of changing variables systematically. So the teacher erased the profit and asked: “How may this be constructed from the other variables?” ($60x4-50$ or $p=ns-k$). Then the profit was reinstated and the selling price was

erased. How might this be found? ($s=(p+k)/n$). After working through each variable separately, the teacher considered variables in pairs. Suppose *both* n and p are erased? How will the profit depend on the number of cards made? Students could then generate a table and/or graph. Finally students might be asked to erase all values and describe the general structure algebraically ($p=ns-k$). This strategy could easily be used whenever students tackle word problems in order to focus more explicitly on structural relationships.

(ii) Assessment Task Genres for Problem Solving

These lessons were designed to assess and improve the capability of students to solve multi-step, non-routine problems and to extend this to the formulation and tackling of problems from the real world. We define a problem as a task that the individual wants to tackle, but for which he or she “does not have access to a straightforward means of solution” (Schoenfeld, 1985). One consequence of this definition is that it is pedagogically inconsistent to design problem-solving tasks for the

Making and Selling Candles	
Original word problem	
A student wants to earn some money by making and selling candles. Suppose that he can make 60 candles from a \$50 kit and that these will each be sold for \$4. How much profit will be made?	
Rewritten problem	
The cost of buying one kit	\$ k 50
The number of candles that can be made with the kit	n 60
The price at which each candle is sold	\$ s 4
Total profit made if all candles are sold.	p \$ 190

Figure 4: Identify and analyze structure: Working with word problems

purpose of practicing a procedure or to develop understanding of a particular concept. In order to develop strategic competence, students must be free to experiment with a range of approaches. They may or may not decide to use any particular procedure or concept; these cannot be pre-determined. Problem solving is contained within the broader processes of mathematical modelling. Modelling additionally requires the formulation of problems by, for example, restricting the number of variables and making simplifying assumptions. Later in the process, solutions must be interpreted and validated in terms of the original context. Some task genres and sample classroom activities for strategic competence are shown in Table 3.

Table 3: Task Genres for Problem Solving Lessons

Assessment task genres	Sample classroom activities.
Solve a non-routine problem by creating an extended chain of reasoning.	Selecting appropriate mathematical concepts and procedures. Planning an approach. Carrying out the plan, monitoring progress and changing direction, where necessary. Reflecting on solutions; examining for reasonableness within the context. Reflecting on strategy; where might it have been improved?
Formulate and interpret a mathematical model of a situation that may be adapted and used in a range of situations.	Making suitable assumptions to simplify a situation. Representing a situation mathematically. Identifying significant variables in situations. Generating relationships between variables. Identifying accessible questions that may be tackled within a situation. Interpreting and validating a model in terms of the context.

The essence of a task in this category is that it should be amenable to a variety of alternative approaches, so that students may learn from comparing these approaches. An example of each type is given in Figure 5. The first is a pure mathematics ‘puzzle’ type problem set in an artificial context, that of a playground game. The second, a modelling task, is taken from a real-life context and involves the student in making simplifications and assumptions. Both however may be tackled in a variety of ways. *The playground game* may be tackled by practical drawing and measuring; by repeated use of Pythagoras’ theorem; and also by ‘pure, non-quantitative, geometric reasoning’. *Having Kittens* may be modelled with a wide variety of representations, and therein is its educational value.

The Playground Game

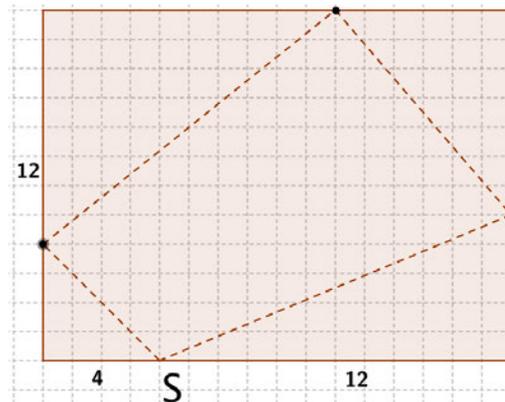
This is a plan view of a 12 meter by 16 meter playground.

The children start at point S, which is 4 meters along the 16-metre wall.

They have to run and touch each of the other three walls and then get back to S.

The first person to return to S is the winner.

What is the shortest route to take?



Having Kittens

Here is a poster published by an organization that looks after stray cats.

**Cats can't add but they do multiply!
In just 18 months, this female cat can have 2000 descendants.**



Figure out whether this number of descendants is realistic.
Here are some facts that you will need:

Length of pregnancy About 2 months	Age at which a female cat can first get pregnant About 4 months	Number of kittens in a litter Usually 4 to 6
Average number of litters a female cat can have in one year 3		Age at which a female cat no longer has kittens About 10 years

Figure 5: Tasks for assessing and improving problem solving processes.

Research-based Principles for Formative Assessment Lesson Design

Having discussed the mathematical focus of the tasks we used, we now turn our attention to how these tasks were incorporated into formative assessment lessons.

The principles that underpinned the design of our lessons were rooted in our “Diagnostic Teaching” program of design research in the 1980s. This was essentially formative assessment under another name (See e.g. Bell, 1993; Swan, 2006a). In a series of studies, on many different topics, we began to define an approach to teaching that we showed were more effective, over the longer term, than either expository or guided discovery approaches (Bassford, 1988; Birks, 1987; Brekke, 1987; Onslow, 1986; Swan, 1983). This approach consisted of four phases. The first involved offering a task designed that would expose students’ existing conceptual understanding and make students aware of their own intuitive interpretations. The second involved the provocation of cognitive conflict by asking students to compare their responses with those of their peers or by asking them to repeat the task using alternative representations and methods. This feedback generated ‘cognitive conflict’ as students began to realize and confront the inconsistencies in their own and each others’ interpretations and methods. Considerable time was then spent reflecting on and discussing the nature of this conflict and students were encouraged to write down the inconsistencies and possible causes of error. The third phase was whole class discussion aimed at resolving conflict. During this phase the teacher would introduce the mathematician’s interpretation. Finally, new learning was ‘consolidated’ by using the newly acquired concepts and methods on further problems. Students were also invited to create and solve their own problems within given constraints, analyze completed work and diagnose causes of error for themselves.

From these studies it was deduced that the value of diagnostic teaching appeared to lie in the extent to which it assessed, identified and focused on the intuitive methods and ideas that students brought to each lesson, and created the opportunity for discussions between students; the greater the intensity of the discussion, the greater was the impact on learning. This is a clear endorsement of the formative assessment practices described in Table 1.

More recently, these results have been replicated on a wider scale. UK government funded the design and development of a multimedia professional development resource to support diagnostic teaching of algebra (Swan & Green, 2002). This was distributed to all Further Education colleges, leading to research on the effects of implementing collaborative approaches to learning in 40 classes of low attaining post 16 students. This again showed the greater effectiveness of approaches that assess and address conceptual difficulties through student-student and whole class discussion (Swan, 2006a, 2006b; Swan, 2006c). A particular design feature of these lessons was the use of a pre- and post-lesson assessment task that would allow both the teacher and the student to assess growth in understanding. The government, recognizing the potential of such resources, commissioned the design of a more substantial multimedia professional development resource, 'Improving Learning in Mathematics' (DfES, 2005). This material was trialed in 90 colleges, before being distributed to all English FE colleges and secondary schools. This material provided many of the resources that were subsequently redeveloped for the Mathematics Assessment Project.

In addition to our own research, we drew inspiration from the ways in which other researchers have structured the design of lessons. These include the Lesson Study research in Japan and the US (Fernandez & Yoshida, 2004; Shimizu, 1999). In Japanese classrooms, lessons are often structured into four phases: *hatsumon* (the teacher gives the class a problem to discuss); *kikan-shido* (the students tackle the problem in groups or individually); *neriage* (a whole class discussion in which alternative strategies are compared and contrasted and in which consensus is sought) and finally the *matome*, or summary, where teachers comment on the qualities of the approaches used. Formative assessment is clearly evident in the way in which the teacher carefully observes students working during the *hatsumon* and *kikan-shido* phases, and selects the ideas to be discussed in the *neriage* stage. The *neriage* phase is considered the most crucial. This term also refers to kneading or polishing in pottery, where different colours are blended together. This serves as a metaphor for the selection and blending of students' ideas. It involves great skill on the part of the teacher, as she must assess student work carefully then select and sequence examples in a way that will elicit fruitful discussions.

Other researchers have adopted similar models for structuring classroom activity. They too emphasize the importance of: anticipating student responses to demanding tasks; carefully monitoring student work; discerning the value of alternative approaches; purposefully selecting ideas for whole class discussion; orchestrating this discussion to build on the collective sense-making of students by careful sequencing of the work to be shared; helping students make connections between and among different approaches and looking for generalizations, and recognizing and valuing and students' constructed solutions by comparing this with existing valued knowledge (Brousseau, 1997; Chazan & Ball, 1999; Lampert, 2001; Stein, Eagle, Smith, & Hughes, 2008).

In order to illustrate how these principles, together with the key strategies in Table 1, have influenced the design of our lessons, we now illustrate the design of complete lessons.

Examples of Formative Assessment Lessons

We now illustrate how this research has informed the lesson structure of the *Classroom Challenges*, integrating the formative assessment strategies of Table 1. A complete lesson guide for this and the other lessons may be downloaded from <http://map.mathshell.org>.

A Concept Development Lesson

The objective of this lesson is to provide a means for a teacher to formatively assess students' capacity to interpret distance-time graphs. The lesson is preceded by a short diagnostic assessment, designed to expose students' prior understandings and interpretations (Figure 6). We encourage teachers to prepare for the lesson by reading through students' responses and by preparing probing

questions that will advance student thinking. They are advised not to score or grade the work. Through our trials of the task, we have developed a “common issues table” that forewarns teachers of some common interpretations students may have, and suggests questions that the teacher might pose to advance a student’s thinking. This form of feedback has been shown to more powerful than grades or scores, which detract from the mathematics and encourage competition rather than collaboration (Black et al., 2003; Black & Wiliam, 1998). Some teachers like to write their questions on the student work while others prepare short lists of questions for the whole class to consider.

Journey to the bus stop
 Every morning Tom walks along a straight road from his home to a bus stop, a distance of 160 meters. The graph shows his journey on one particular day.

Time (seconds)	Distance from home (meters)
0	0
10	20
20	40
30	60
40	80
50	100
60	60
70	40
80	80
90	120
100	160
110	160
120	160

- Describe what may have happened. Include details like how fast he walked.
- Are all sections of the graph realistic? Fully explain your answer.

<i>Issue</i>	<i>Suggested questions and prompts</i>
<p>Student interprets the graph as a picture</p> <p>For example: The student assumes that as the graph goes up and down, Tom’s path is going up and down <i>or</i> assumes that a straight line on a graph means that the motion is along a straight path.</p>	<p>If a person walked in a circle around their home, what would the graph look like?</p> <p>If a person walked at a steady speed up and down a hill, directly away from home, what would the graph look like?</p> <p>In each section of his journey, is Tom’s speed steady or is it changing?</p> <p>How do you know?</p> <p>How can you figure out Tom’s speed in each section of the journey?</p>
<p>Student interprets graph as speed–time</p> <p>The student has interpreted a positive slope as ‘speeding up’ and a negative slope as ‘slowing down’.</p>	<p>If a person walked for a mile at a steady speed, away from home, then turned round and walked back home at the same steady speed, what would the graph look like?</p> <p>How does the distance change during the second section of Tom’s journey?</p> <p>What does this mean?</p> <p>How can you tell if Tom is traveling away from or towards home?</p>

Figure 6: Initial assessment task: *Journey to school*, and an extract from the ‘Common issues table’

The lesson itself is structured in five parts:

- Make existing concepts and methods explicit.** An initial task is offered with the purpose of clarifying the learning intentions, making students aware of their own intuitive interpretations, creating curiosity and modeling the level of reasoning to be expected during the main activity (Table 1, strategy 1). The teacher displays the task shown in Figure 7 and

Bartell, T. G., Bieda, K. N., Putnam, R. T., Bradfield, K., & Dominguez, H. (Eds.). (2015). *Proceedings of the 37th annual meeting of the North American Chapter of the International Group for the Psychology of Mathematics Education*. East Lansing, MI: Michigan State University.

asks students to select the story that best fits the graph. This usually results in a spread of student opinions, with many choosing option B. The teacher invites and probes explanations, and labels the diagram with these explanations, but does not correct students, nor attempt to reach resolution at this point.

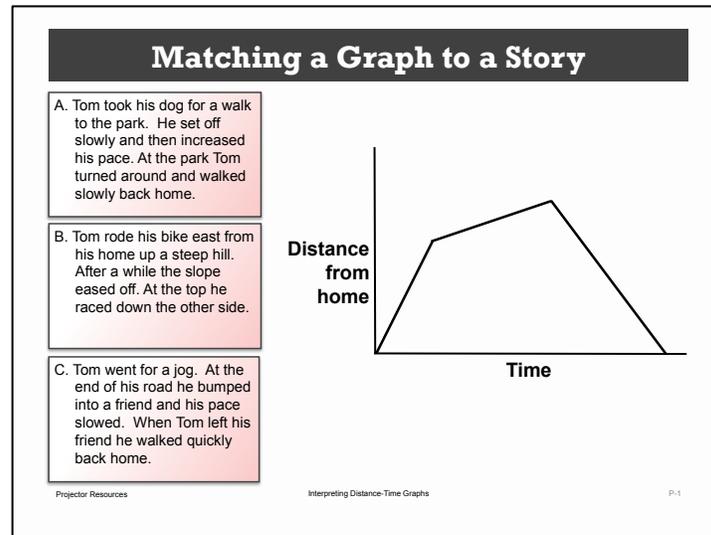


Figure 7: Introductory activity: Interpreting distance-time graphs

2. **Collaborative activity: Matching graphs, stories and tables.**

This phase is designed to create student-student discussions in which they share and challenge each others' interpretations (Table 1, strategy 2). Each group of students is given a set of the cards shown in Figure 8. Ten distance/time graphs are to be matched with nine 'stories' (the tenth to be constructed by the student). Subsequently, when the cards have been discussed and matched, the teacher distributes a further set of cards that contain distance/time tables of numerical data. These provide feedback by enabling students to check their own responses (by plotting if necessary), and reconsider the decisions that have been made. Students collaborate to construct posters displaying their reasoning. While students work, the teacher is encouraged to ask the pre-prepared questions from the initial diagnostic assessment (Table 1, strategy 3).

3. **Inter-group discussion: Comparing interpretations.** Students' posters are displayed, and students visit each other's posters and check them, demanding explanations for matches that do not appear to be correct (Table 1, strategy 4).
4. **Plenary discussion.** Students revisit the task that was introduced at the beginning of the lesson and resolution is now sought. Drawing on examples of student work produced during the lesson, the teacher draws attention to the significant concepts that have arisen (e.g. the connection between speed, slopes on graphs, and differences in tables). Further questions are posed to check learning, using mini-whiteboards. "Show me a distance time graph to show this story"; "Show me a story for this graph"; "Show me a table that would fit this graph". (Table 1, strategy 2)
5. **Individual work: Improving solutions to the pre-assessment task.** Students now revisit the work they did on the pre-assessment task. They describe how they would now answer the task differently and write about what they have learned. They are also asked to solve a fresh, similar task (Table 1, strategy 5).

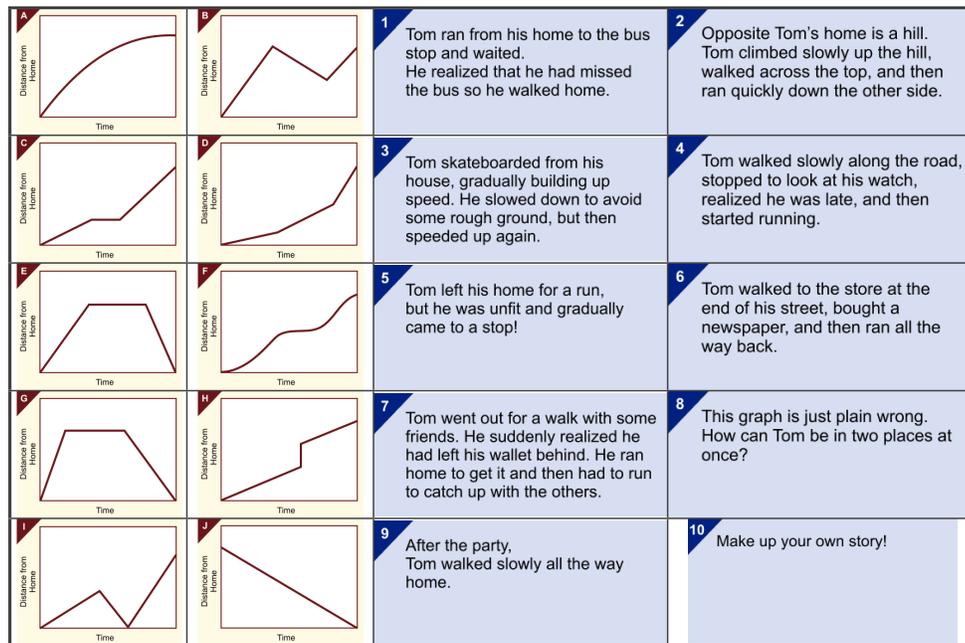


Figure 8: Matching cards: Graphs and stories.

A Problem Solving Lesson

The problem solving lessons were constructed in a similar way, but with a different emphasis. Teachers found it very difficult to interpret, monitor and select students' extended reasoning during a problem-solving lesson. We therefore decided again to precede each lesson with a preliminary assessment in which students tackle the problem individually. The teacher reviews a sample of the students' initial attempts and identifies the main issues that need addressing. This time the issues focus on *approaches* to the problem. If time permits, teachers write feedback questions on each student's work, or alternatively prepare questions for the whole class to consider. Figure 9 illustrates some of the common issues and suggested questions for the task "Having Kittens" (Figure 5).

<i>Issue</i>	<i>Suggested questions and prompts</i>
Has difficulty starting	Can you describe what happens during first five months?
Does not develop a suitable representation	Can you make a diagram or table to show what is happening?
Work is unsystematic	Could you start by just looking at the litters from the first cat? What would you do after that?
Develops a partial model	Do you think the first litter of kittens will have time to grow and have litters of their own? What about their kittens?
Does not make clear or reasonable assumptions	What assumptions have you made? Are all your kittens born at the beginning of the year? Are all your kittens females?
Makes a successful attempt	How could you check this answer using a different method?

Figure 9: An extract from the 'Common issues table' for *Having Kittens*

Now we come to the lesson itself. While the precise structure is problem-specific, these lessons are generally structured as follows:

Bartell, T. G., Bieda, K. N., Putnam, R. T., Bradfield, K., & Dominguez, H. (Eds.). (2015). *Proceedings of the 37th annual meeting of the North American Chapter of the International Group for the Psychology of Mathematics Education*. East Lansing, MI: Michigan State University.

1. **Introduction.** The teacher re-introduces the main task for the lesson and returns students' work along with the formative questions. Students are given a few minutes to read these questions and respond to them, individually (Table 1, strategy 3)
2. **Group work: comparing strategic approaches.** The students are asked to work in small groups to discuss the work of each individual, then to produce a poster showing a joint solution that is better than the individual attempts. Groups are organised so that students with contrasting ideas are paired. This activity promotes peer assessment and collaboration. The teacher's role is to observe groups and challenge students using the prepared questions and thus refine and improve their strategies (Table 1, strategy 2).
3. **Inter-group discussion: comparing strategic approaches.** Depending on the range of approaches in evidence, the teacher may at this point ask students to review the strategic approaches produced by other groups in the class, and justify their own. (Most will not have arrived at a solution by this stage). If there is not a sufficient divergence of methods, or more sophisticated representations are not becoming apparent, then the teacher may move directly to the next stage. (Table 1, strategy 4)
4. **Group work: critiquing pre-designed 'sample student work'.** The teacher introduces up to four pieces of "sample student work", provided in the materials (Figure 10). This work has been chosen to highlight significant, alternative approaches. For example, it may show different representations of the situation. Each piece of work is annotated with questions that focus students' attention. (E.g. "What has each student done correctly? What assumptions have they made? How can their work be improved?") This intervention is discussed further in the following section.
5. **Group work: refining solutions.** Students are given an opportunity to respond to the review of approaches. They revisit the task and try to use insights to further refine their solution (Table 1, strategy 4).
6. **Whole class discussion: a review of learning.** The teacher holds a plenary discussion to focus on the processes involved in the problem, such as the implications of making different assumptions, the power of alternative representations and the general mathematical structure of the problem. This may also involve further references to the approaches in the sample student work.

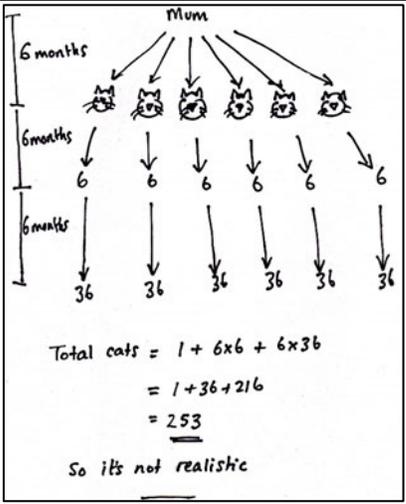
 <p style="text-align: center;"> $\text{Total cats} = 1 + 6 \times 6 + 6 \times 36$ $= 1 + 36 + 216$ $= \underline{253}$ </p> <p style="text-align: center;">So its not realistic</p>	<p>Questions for students</p> <ul style="list-style-type: none"> • What has Wayne done correctly? • What assumptions has he made? • How can Wayne's work be improved? <p>Notes from the teacher guide</p> <p>Wayne has assumed that the mother has six kittens after 6 months, and has considered succeeding generations. He has, however, forgotten that each cat may have more than one litter. He has shown the timeline clearly. Wayne doesn't explain where the 6-month gaps have come from.</p>
---	--

Figure 10: Sample work for discussion, with commentary from the teacher guide.

The above lesson description contains many features that are not common in mathematics teaching, at least in the US and UK. There is a strong emphasis on the use of preliminary formative assessment, which enables the teacher to prepare for and adapt interventions to the student reasoning that will be encountered. Students spend much of the lesson in dialogic talk, focused on comparing mathematical processes. The successive opportunities for refining the solution enable students to pursue multiple methods, and to compare and evaluate them. Finally, designed ‘sample student work’ is used to foster the development of *critical competence*. This aspect has become the focus of our recent research, and we now draw out some of the issues this raises.

Students Assessing Student Work

In Cobb’s terms, the products of design research are ‘humble’ theories that guide future designs (Cobb et al., 2003). As we have worked through successive refinements, many of the findings from the data have been incorporated into the designs themselves. Below we just one of the features of these lessons that we are continuing to study further (Evans & Swan, 2014); that of students critiquing pre-designed ‘sample student work’.

Researchers (e.g. Stein et al., 2008) have emphasised the importance of students assessing approaches to cognitively demanding tasks, but this has proved difficult for teachers to put into practice, particularly for problem solving, where student reasoning is extended, complex and often poorly articulated. In a busy classroom, teachers find it difficult to observe, interpret and select suitable work for sharing. In whole class discussions we frequently observe students presenting posters of their reasoning, to a sea of incomprehension. Teachers also find it difficult to quickly recognize and make connections between students’ ideas and draw out significant learning points. It is therefore understandable that, in practice, the sharing of ideas often degenerates into mere ‘show and tell’, with participation prioritized over learning (Stein et al., 2008).

In response to this challenge we are researching the potential uses of pre-designed ‘sample student work’ to focus classroom discussion on key concepts and processes, while at the same time developing critical competence. We construct this work by analyzing a sample of genuine student responses to a problem, then identifying conceptual difficulties or problem solving strategies that will provide significant learning opportunities for students. When problem solving, for example, very few students autonomously decide to employ an algebraic method (Treilibs, 1979). Given choice they tend to resort to more secure numerical or graphical methods. For this reason we may include an algebraic method among the sample work so that students will be confronted with methods they may not yet have considered. We present this work in clear, legible, handwritten form, to suggest that the work is tentative, open for criticism and improvement. We have found that students feel more able to criticize such work than the work of peers, where social pressures often come into play.

We have found that pre-designed sample student work has many potential uses. In problem solving, for example, it can be used to encourage a student that is stuck in one line of thinking to consider others, to enable comparison of alternative representations and to focus on the identification of modeling assumptions. In concept learning it may be used to draw attention to common mathematical misconceptions and alternative interpretations. Perhaps most importantly, the sample work may provide an opportunity for ‘clarifying our learning intentions and criteria for success’ (Table 1, strategy 1). By assessing the work of others, students become more aware of the criteria by which their own work is judged. Thus, for example, by asking students to compare four methods and judge which is most ‘powerful’, ‘clear’, or ‘elegant’, then they may come to understand what such terms may mean.

In our classroom observations (in the UK and the US), however, we found that there were frequent problems with implementation (Evans & Swan, 2014). These included: students commenting superficially, focusing merely on presentation and clarity; students being given

insufficient time to engage with the reasoning presented in the work; students spending time correcting errors rather than focusing on strategy; students not using the work to improve their own solutions; students failing to make comparisons between approaches. In response, we established the following guidelines for the design of sample work:

- Discourage superficial analysis by students, by stating explicitly the purpose of the sample work, and by asking specific questions that relate to this purpose;
- Encourage holistic comparisons by making the sample work short, accessible and clear, and by excluding procedural and other errors that distract attention away from the identified purpose;
- Leave the work unfinished, so that students have to engage with the reasoning in order to complete it;
- Sequence the distribution of the sample student work so that successive pairwise comparisons of approaches may be made;
- Offer students sufficient time and opportunity to incorporate what they have learned from the sample work into their own solutions;
- Offer the teachers support for the whole class discussion so that they can identify and draw out criteria for the comparison of alternative approaches.

When these guidelines were followed, however, we found that critiquing work provides the potential to refocus students' attention away from 'getting answers' towards 'thinking about reasoning' and a deeper awareness of the learning intentions of the teacher and the criteria for success.

Concluding Remarks

In this brief paper, I have attempted to describe how systematic design research has enabled us to tackle a significant pedagogical problem: how might we enable teachers to embed formative assessment practices into their normal classroom practice? I have discussed the five strategies described by Black and Wiliam and shown how these have been integrated into the structure of the *Classroom Challenges*. In particular, I have attempted to show how:

- Learning intentions and criteria for success may be clarified by making use of task genres that *require* the mathematical practices that we seek to foster; by sharing these intentions and modeling the reasoning required at the beginning of lessons; and by encouraging students to focus on criteria for success as they critique and evaluate the work of others.
- Evidence of student understanding may be elicited through: pre-assessment tasks that offer students opportunity to engage with a problem individually, before group discussion takes place; and through group activities that require shared resources and dialogic talk in which students share interpretations and strategies. These give the teacher opportunities to reflect on student reasoning and to plan and make appropriate interventions.
- Common issues tables may be used to help teachers plan appropriate feedback that will prompt students to reconsider their thinking and move them forward.
- Students may become instructional resources for one another as they work collaboratively and review and comment on the work of their peers.
- Students may take a greater responsibility for their own learning as they become more aware of what they have learned and what they still need to learn through reflection at the end of lessons and through the matching of their own responses to the designed sample student work.

Of course, we realize that however carefully we design lesson structures, each classroom is unique and teachers will modify what we offer in their own way. Early evidence of their impact is, however, encouraging. Drawing on a national survey of 1239 mathematics teachers from 21 states, and interview data from four sites, Research for Action (RFA, 2015), found that a large majority of teachers reported that the use of the *Classroom Challenges* had helped them to implement the Common Core State Standards, raise their expectations for students, learn new strategies for teaching subject matter, use formative assessment; and differentiate instruction.

The National Center for Research on Evaluation, Standards and Student Testing (CRESST) examined the implementation and impact of *Classroom Challenges* in 9th grade Algebra 1 classes (Herman et al., 2015). This study used a quasi-experimental design to compare student performance with *Classroom Challenges* to a matched sample of students from across Kentucky comparable in prior achievement and demographic characteristics. On average, study teachers implemented only four to six *Challenges* during the study year (or 8-12 days), yet, relative to typical growth in math from eighth to ninth grade, the effect size for the *Classroom Challenges* represented an additional 4.6 months of schooling. Although teachers felt that the *Challenges* benefited students' conceptual understanding and mathematical thinking, they reported that sizeable proportions of their students struggled, and it appeared that lower achieving students benefitted less than higher achievers. This they suggested, may be due to the great difference in challenge and learning style required by these lessons, compared with their previous diet of procedural learning.

Finally, Inverness Research (IR, 2015) in 2014 surveyed 636 students from 31 trial classes (6th grade to High School) across five states in the US. They found that the majority of students enjoyed learning math through these lessons, reported that they understood it better, had increased in their participation, listening to others, and in explaining their mathematical thinking. About 20%, however, remained unaffected by or disaffected with these lessons. This was because they didn't enjoy working in groups, they objected to the investigative approach, and/or they felt that these lessons were too long, or too difficult

In conclusion, it does appear that the *Classroom Challenges* provide a model for teachers as they attempt to introduce formative assessment into their everyday classroom practice, but they do require a radical shift in the predominant culture within many classrooms. The potential for improving learning through the integration of these formative assessment practices into everyday teaching is, however, clear. This project has shown that classroom materials with this focus can help teachers make it a reality in their classrooms. How far teachers transfer this approach into the rest of their teaching is the focus of ongoing research.

Acknowledgements

The Mathematics Assessment Program (MAP) materials were produced as part of a collaboration between the University of California, Berkeley and the Shell Center team at the University of Nottingham, with support from the Bill & Melinda Gates Foundation. The team works with the Silicon Valley Mathematics Initiative and school systems across the US and UK to develop improved assessment.

The Having Kittens task was originally designed by Acumina Ltd. (<http://www.acumina.co.uk/>) for Bowland Maths (<http://www.bowlandmaths.org.uk>) and appears courtesy of the Bowland Charitable Trust.

References

- Barab, S., & Squire, K. (2004). Design-based research: Putting a stake in the ground. *The Journal of the Learning Sciences*, 13(1), 1-14.
- Bassford, D. (1988). *Fractions: A comparison of Teaching Methods*. (M.Phil), University of Nottingham.
- Bell, A. (1993). Some experiments in diagnostic teaching. *Educational Studies in Mathematics*, 24(1).

Bartell, T. G., Bieda, K. N., Putnam, R. T., Bradfield, K., & Dominguez, H. (Eds.). (2015). *Proceedings of the 37th annual meeting of the North American Chapter of the International Group for the Psychology of Mathematics Education*. East Lansing, MI: Michigan State University.

- Bereiter, C. (2002). Design research for sustained innovation. *Cognitive studies, Bulletin of the Japanese Cognitive Science Society*, 9(3), 321-327.
- Birks, D. (1987). *Reflections: a Diagnostic Teaching Experiment*. University of Nottingham.
- Black, P., Harrison, C., Lee, C., Marshall, B., & Wiliam, D. (2003). *Assessment for learning: Putting it into practice*. Buckingham: Open University Press.
- Black, P., & Wiliam, D. (1998). *Inside the black box: raising standards through classroom assessment*. London: King's College London School of Education 1998.
- Black, P., & Wiliam, D. (1999). *Assessment for learning: beyond the black box*. Cambridge: University of Cambridge Institute of Education.
- Black, P., & Wiliam, D. (2009). Developing the theory of formative assessment. *Educational Assessment, Evaluation and Accountability*, 21(1), 5-31.
- Brekke, G. (1987). *Graphical Interpretation: a study of pupils' understanding and some teaching comparisons*. University of Nottingham.
- Brousseau, G. (1997). *Theory of Didactical Situations in Mathematics* (N. Balacheff, M. Cooper, R. Sutherland, & V. Warfield, Trans. Vol. 19). Dordrecht: Kluwer.
- Burkhardt, H. (2006). From design research to large scale impact. In J. van den Akker, K. Gravemeijer, S. McKenney, & N. Nieveen (Eds.), *Educational Design Research* (pp. 121-150). Abingdon, Oxon: Routledge.
- Chazan, D., & Ball, D. L. (1999). Beyond being told not to tell. *For the Learning of Mathematics*, 19(2), 2-10.
- Cobb, P., Confrey, J., diSessa, A., Lehrer, R., & Schauble, L. (2003). Design Experiments in Educational Research. *Educational Researcher*, 32(1).
- DBRC. (2003). Design-based research: An emerging paradigm for educational inquiry. *Educational Researcher*, 32(1), 5-8.
- DfES. (2005). *Improving Learning in Mathematics*. London: Standards Unit, Teaching and Learning Division.
- Evans, S., & Swan, M. (2014). Developing students' strategies for problem solving in mathematics: the role of pre-designed "Sample Student Work". *Educational Designer*, 2(7).
- Fernandez, C., & Yoshida, M. (2004). *Lesson Study: A Japanese Approach to Improving Mathematics Teaching and Learning* Mahwah, New Jersey Laurence Erlbaum Associates.
- Hatano, G., & Inagaki, K. (1986). Two courses of expertise. In H. W. Stevenson, H. Azuma, & K. Hakuta (Eds.), *Child development and education in Japan* (pp. 262-272). New York: W H Freeman/Times Books/ Henry Holt & Co.
- Herman, J., Epstein, S., Leon, S., La Torre Matrundola, D., Reber, S., & Choi, K. (2015). Implementation and Effects of LDC and MDC in Kentucky Districts. *CRESST Policy Brief No. 13*. Retrieved from: http://www.cse.ucla.edu/products/policy/PB_13.pdf
- IR. (2015). Mathematics Assessment Program (MAP): Project Portfolio. from http://inverness-research.org/mars_map/xindex.html
- ISDDE. (2012). High-stakes Examinations to Support Policy. *Educational Designer*, 2(5).
- Jones, I., Pollitt, A., & Swan, M. (2015). Assessing mathematical problem solving using comparative judgement. *International Journal of Science and Mathematics Education*, 13(1), 151-177.
- Kelly, A. (2003). Theme issue: The role of design in educational research. *Educational Researcher*, 32(1), 3-4.
- Lakatos, I. (1976). *Proofs and Refutations*. Cambridge: Cambridge University Press.
- Lampert, M. (2001). *Teaching problems and the problems of teaching*. New Haven, CT: Yale University Press.
- NGA, & CCSSO. (2010). Common Core State Standards for Mathematics: National Governors Association, Council of Chief State School Officers.
- Onslow, B. (1986). *Overcoming conceptual obstacles concerning rates: Design and Implementation of a diagnostic Teaching Unit*. (PhD), University of Nottingham.
- RFA. (2015). MDC's Influence on Teaching and Learning. from <http://www.researchforaction.org/rfa-study-of-tools-aligned-ccss/>
- Schoenfeld, A. H. (1985). *Mathematical Problem Solving*: Academic Press.
- Shimizu, Y. (1999). Aspects of Mathematics TEacher Education in Japan: Focusing on Teachers' Roles. *Journal of Mathematics Teacher Education*, 2, 107-116.
- Sierpinska, A. (1994). *Understanding in Mathematics*, . London: Falmer.
- Stein, M. K., Eagle, R. A., Smith, M. A., & Hughes, E. K. (2008). Orchestrating productive mathematical discussions: Five practices for helping teachers move beyond show and tell. *Mathematical Thinking and Learning*, 10, 313-340.

Bartell, T. G., Bieda, K. N., Putnam, R. T., Bradfield, K., & Dominguez, H. (Eds.). (2015). *Proceedings of the 37th annual meeting of the North American Chapter of the International Group for the Psychology of Mathematics Education*. East Lansing, MI: Michigan State University.

- Swan, M. (1983). *Teaching Decimal Place Value - a comparative study of 'conflict' and 'positive only' approaches*. Paper presented at the 7th Conference of International Group for the Psychology of Mathematics Education, Jerusalem, Israel.
- Swan, M. (2006a). *Collaborative Learning in Mathematics: A Challenge to our Beliefs and Practices*. London: National Institute for Advanced and Continuing Education (NIACE) for the National Research and Development Centre for Adult Literacy and Numeracy (NRDC).
- Swan, M. (2006b). Designing and using research instruments to describe the beliefs and practices of mathematics teachers. *Research In Education*, 75, 58-70.
- Swan, M., & Green, M. (2002). *Learning Mathematics through Discussion and Reflection*. London: Learning and Skills Development Agency.
- Swan, M. B. (2006c). Learning GCSE mathematics through discussion: what are the effects on students? *Journal of Further and Higher Education*, 30(3), 229-241. doi: 10.1080/03098770600802263
- Treilibs, V. (1979). *Formulation Processes in Mathematical Modelling*. (MPhil), University of Nottingham, Nottingham.
- van den Akker, J., Graveemeijer, K., McKenney, S., & Nieveen, N. (Eds.). (2006). *Educational Design Research*. London and New York: Routledge.
- Wiliam, D., & Thompson, M. (2007). Integrating assessment with instruction: What will it take to make it work? . In C. A. Dwyer (Ed.), *The future of assessment: shaping teaching and learning* (pp. 53-82). Mahwah, NJ: Lawrence Erlbaum Associates.
- Zaslavsky, O. (2008). *Attention to similarities and differences: A fundamental principle for task design and implementation in mathematics education*. . Paper presented at the International Congress in Mathematics Education, Mexico. <http://tsg.icme11.org/tsg/show/35>