



## Foster, Colin (2016) Confidence and competence with mathematical procedures. *Educational Studies in Mathematics*, 91 (2). pp. 271-288. ISSN 1573-0816

### Access from the University of Nottingham repository:

<http://eprints.nottingham.ac.uk/30544/1/Foster%2C%20Confidence%20and%20competence%20with%20mathematical%20procedures.pdf>

### Copyright and reuse:

The Nottingham ePrints service makes this work by researchers of the University of Nottingham available open access under the following conditions.

- Copyright and all moral rights to the version of the paper presented here belong to the individual author(s) and/or other copyright owners.
- To the extent reasonable and practicable the material made available in Nottingham ePrints has been checked for eligibility before being made available.
- Copies of full items can be used for personal research or study, educational, or not-for-profit purposes without prior permission or charge provided that the authors, title and full bibliographic details are credited, a hyperlink and/or URL is given for the original metadata page and the content is not changed in any way.
- Quotations or similar reproductions must be sufficiently acknowledged.

Please see our full end user licence at:

[http://eprints.nottingham.ac.uk/end\\_user\\_agreement.pdf](http://eprints.nottingham.ac.uk/end_user_agreement.pdf)

### A note on versions:

The version presented here may differ from the published version or from the version of record. If you wish to cite this item you are advised to consult the publisher's version. Please see the repository url above for details on accessing the published version and note that access may require a subscription.

For more information, please contact [eprints@nottingham.ac.uk](mailto:eprints@nottingham.ac.uk)

# **Confidence and Competence with Mathematical Procedures**

Colin Foster

colin.foster@nottingham.ac.uk

## **Abstract**

Confidence assessment (CA), in which students state alongside each of their answers a confidence level expressing how certain they are, has been employed successfully within higher education. However, it has not been widely explored with school pupils. This study examined how school mathematics pupils ( $N = 345$ ) in five different secondary schools in England responded to the use of a CA instrument designed to incentivise the eliciting of truthful confidence ratings in the topic of directed (positive and negative) numbers. Pupils readily understood the negative marking aspect of the CA process and their facility correlated with their mean confidence with  $r = .546$ ,  $N = 336$ ,  $p < .001$ , indicating that pupils were generally well calibrated. Pupils' comments indicated that the vast majority were positive about the CA approach, despite its dramatic differences from more usual assessment practices in UK schools. Some pupils felt that CA promoted deeper thinking, increased their confidence and had a potential role to play in classroom formative assessment.

## **Keywords**

certainty-based assessment; confidence-based assessment; directed numbers; formative assessment; mathematics education; mixed methods research; negative numbers

## **1. Introduction**

Fluency in important mathematical procedures is now recognised as a critical goal within the learning of school mathematics (Gardiner, 2014; NCTM, 2014). Rather than constituting a threat to conceptual understanding, procedural fluency is its natural partner (Foster, 2013; Hewitt, 1996; Kent & Foster, 2015; Kieran, 2013). Indeed, procedural and conceptual learning are now increasingly seen to be interrelated and inseparable (Baroody, Feil, & Johnson, 2007; Star, 2005, 2007), since security with fundamental procedures offers pupils increased power to explore more complicated mathematics at a conceptual level (Foster, 2014). East Asian countries with impressive performances in large-scale international assessments such as the Programme for International Student Assessment (PISA) and the Trends in International Mathematics and Science Study (TIMSS) succeed in emphasising the development of mathematical fluency without resorting to the low-level rote learning of procedures (Askew, Hodgen, Hossain, & Bretscher, 2010; Fan & Bokhove, 2014; Leung, 2014). Finding ways to support the meaningful learning of mathematical procedures and the development of robust fluency with important mathematical skills is an urgent priority for Western mathematics education. The new national curriculum for mathematics in England, for example, emphasises procedural fluency as the first stated aim (Department for Education, 2013) and provides a political imperative in the UK for developing this important aspect of learning mathematics.

Procedural fluency involves knowing when and how to apply a procedure and being able to perform it “accurately, efficiently, and flexibly” (NCTM, 2014, p. 1). Written mathematics assessments generally focus on competence with procedures, but an important concomitant of procedural competence is the pupil’s confidence in the answers that they give (Kyriacou, 2005). If a pupil’s performance of a mathematical procedure does not result in an answer that the pupil believes is very likely to be correct, then it is not a useful tool for them; under such

circumstances it would be better to approach the problem by some other method, seek help or use a calculator or computer to give a reliable answer. Thus for secure development of procedural fluency it is important not only that a pupil can obtain the correct answer in a reasonable amount of time but that they have an accurate sense of their reliability with the procedure.

It is common in preparation for summative assessments for school pupils to be instructed to guess any answers of which they are unsure. This may be presented as an aspect of “examination technique”, since leaving a blank response is guaranteed to result in the award of no marks, so any answer is better than none (Foster, 2007). However, in any practical circumstance beyond an educational setting, a person’s knowing whether they know is extremely important, as it determines whether additional support (from other people, a computer or a reference source) is required. For example, it is much better to know that you do not know the value of  $-2 + 7$ , although you think that it might be  $-5$ , than it is to be sure that it is  $-5$  when in fact it is 5. In the classroom too, it is clearly extremely valuable for the teacher to know whether a pupil’s incorrect response is a wild guess or whether it might constitute evidence of a deep-seated misunderstanding or misconception. Thus knowing how strongly the pupil believes in the answer that they are giving can be potentially very helpful formative assessment for the teacher in judging what subsequent learning activities or interventions might be most appropriate.

This important kind of confidence is unlikely to be captured by pupils’ global judgments of their feelings of confidence for mathematics or particular topics within the mathematics curriculum. Much research has been carried out in the field of affect, such as explorations into the interplay between cognition, beliefs and attitudes (Di Martino & Zan, 2011; Hannula, 2011; Pepin & Roesken-Winter, 2014). However, broad constructs at the subject level, such as “mathematics confidence” (Pierce & Stacey, 2004), are too general for the purpose here.

Even a topic-level perception of confidence, such as “I am good at negative numbers”, is of limited usefulness when faced with a specific calculation such as  $-2 + 7$ . Within any particular topic, competence may vary significantly at the level of detail of particular questions, so it is important to explore pupils’ confidence at a similarly fine-grained level. Pupils whose confidence and competence are strongly correlated are said to be “well calibrated”:

An individual is well calibrated if, over the long run, for all propositions assigned a given probability, the proportion that is true is equal to the probability assigned. For example, half of those statements assigned a probability of .50 of being true should be true, as should 60% of those assigned .60, and all of those about which the individual is 100% certain. (Fischhoff, Slovic, & Lichtenstein, 1977, p. 552)

It follows that in order to remain well calibrated, a pupil whose competence varies across some domain needs also to have varying confidence levels across that same domain.

In this study I examine the relationship between the confidence and competence of school pupils within an important secondary mathematics topic: directed (positive and negative) numbers. I explore how well calibrated pupils are within this topic by using an instrument designed to incentivise the eliciting of truthful confidence ratings. I examine pupils’ comments on the use of this instrument and discuss what roles such tools might be able to play in formative and summative assessment within school mathematics.

## **2. Confidence and competence**

Within the mathematics education literature, confidence has been construed in different ways. The Confidence in Learning Mathematics Scale within the Fennema-Sherman Mathematics Attitudes Scales (MAS) (Fennema & Sherman, 1976) has been used extensively to study pupils’ attitudes towards mathematics (Lim & Chapman, 2013). Taking a social perspective,

Burton (2004, p. 360) saw confidence as “a label for a confluence of feelings relating to beliefs about the self, and about one’s efficacy to act within a social setting, in this case the mathematics classroom”. More specifically, Galbraith and Haines (1998) stated:

Students with high mathematics confidence believe they obtain value for effort, do not worry about learning hard topics, expect to get good results, and feel good about mathematics as a subject. Students with low confidence are nervous about learning new material, expect that all mathematics will be difficult, feel that they are naturally weak at mathematics, and worry more about mathematics than any other subject. (p. 278)

Similar to this, Pierce and Stacey (2004, p. 290) defined mathematics confidence, as “a student’s perception of their ability to attain good results and their assurance that they can handle difficulties in mathematics”. Some of these notions of confidence encompass mathematical self-efficacy, which is a pupil’s belief in advance about their likelihood of successfully performing a particular mathematical task. Bandura (1977) argued that a person’s self-efficacy is a major factor in whether they will attempt a given task, how much effort they will be prepared to put in, and how resilient they will be when difficulties arise. More recently the construct of mathematical resilience has been developed to capture the ways in which pupils overcome barriers to learning mathematics (Johnston-Wilder and Lee, 2010).

Many factors, both within and outside the classroom, are likely to be important in affecting a pupil’s mathematical self-confidence. A meta-analysis of studies on gender differences in mathematics (Frost, Hyde, & Fennema, 1994) found that girls had lower mathematics self-concept/confidence than boys and greater mathematics anxiety, although the effect sizes were small. Girls’ self-confidence is lower in general (Hannula, et al., 2005; Leder, 1995), and there is evidence that this effect is more marked in higher-attaining sets (Jones, 1995). In a

case study of two schools, Boaler (1998) found that at the school where pupils were taught mathematics in attainment sets and used a traditional, textbook approach, boys showed greater confidence than girls. However, at the school which used all-attainment classes and open-ended activities there was no difference between boys' and girls' confidence. It is likely that multiple factors are important, so drawing simplistic conclusions from this one study should be avoided.

However, in contrast to research in which confidence is viewed broadly as an overall perception of learning mathematics, the focus in this paper is on pupils' confidence with regard to their responses to specific items. Whereas mathematical self-efficacy (Bandura, 1977) involves a pupil's anticipatory judgment about their likelihood of successfully answering a particular mathematical item, in this study pupils were invited to ascribe a confidence level after the item was completed. This is quite different, since at this point the mathematics called on, and its demands, are known, and the reasonableness of the answer can also be assessed (Morony, Kleitman, Lee, & Stankov, 2013). Stankov, Lee, Luo and Hogan (2012) defined confidence as "a state of being certain about the success of a particular behavioral act" (p. 747), and in this study we consider a pupil's "confidence of response" as how certain they are that the answer that they have just given is correct. Unlike self-efficacy, this kind of confidence is after-the-fact and can even be ascribed to answers produced by another pupil.

It is natural to see the variables confidence and competence as defining a two-dimensional space (Figure 1). Well-calibrated pupils (Fischhoff, Slovic, & Lichtenstein, 1977), whose confidence closely matches their competence, would be located near to the diagonal line in Figure 1, with over-confident and under-confident pupils taking positions on either side. Viewed in this way, traditional assessments which measure only competence give a partial and perhaps misleading picture. Gattegno (1987) commented on the difference between a

pupil answering the question “ $2 + 3$ ” with a querying intonation “Five?” as opposed to a more declamatory “Five!”, interpreting these as manifestations of differing levels of confidence that could require quite different teacher responses. Both for summative and for formative purposes, it may be extremely important to know where pupils are positioned horizontally in Figure 1, as well as vertically.

Pupils’ responses in the classroom (both oral and written) can range from a wild guess to an assured answer, and the level of confidence has a potential role to play within formative assessment (Black & Wiliam, 1998; Black et al., 2003; Black & Wiliam, 2009; Clark, 2015; Warwick, Shaw, & Johnson, 2015). Gardner-Medwin and Gahan (2003) described students with different levels of belief regarding a true statement as being in a state of “knowledge, uncertainty, ignorance, misconception [or] delusion”, commenting that ignorance is “far from the worst state to be in” (p. 147). The teacher might choose to intervene in quite different ways with two pupils with similar competence but very different confidence. A pupil with high competence but low confidence (represented by the dot in the top left of Figure 1) might be encouraged to return to ideas with which they are confident and build from there, whereas a pupil with high but misplaced confidence (represented by the dot in the bottom right of Figure 1) might benefit from experiencing some cognitive conflict to challenge relevant misconceptions (see the arrows in Figure 1). In both cases, movement towards better calibration is desired, but this might be effected in quite different ways.



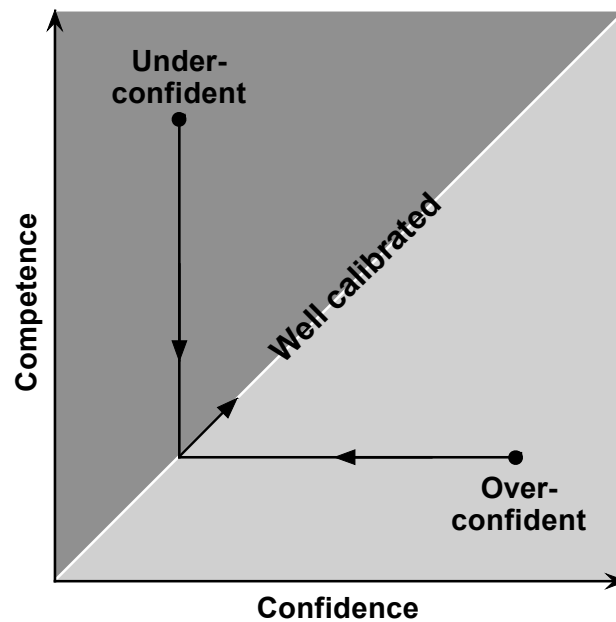


Figure 1. Two pupils and possible trajectories towards better calibration.

### 3. Confidence assessment

Confidence is difficult to assess reliably, since pupils may be inclined to exaggerate their confidence in order to win the approval of the teacher (or fend off unwanted attention) or raise their status in the eyes of their peers (Hannula, 2003). The common classroom practices of asking pupils to indicate by raising their hands whether they are sure (yes or no) or to specify their level of confidence using traffic lights (red/amber/green) both suffer from this problem of over-reporting. Inviting pupils to explain their reasoning can be a powerful technique for formative assessment of confidence, but may become repetitive and unengaging for the rest of the class – and explanations can be rote learned as well as procedures (Kent & Foster, 2015).

Confidence-based assessment (CA) is one practical solution to the problem of accurately measuring pupils' confidence, and this now plays an important role in higher education in the study of medicine and related disciplines (Gardner-Medwin, 1995, 1998, 2006; Schoendorfer & Emmett, 2012). It is easy to see the need for medical practitioners to recognise when they

are uncertain of a course of action, as mistakes can be costly in human terms. Such assessment practices seek to reward self-awareness by giving marks for confidence in correct responses while subtracting marks for misplaced confidence in incorrect responses. Typically, students are asked to give a confidence level of 1, 2 or 3 for each of their responses. If their response is correct, the number of marks that they receive for that question is the same as their chosen confidence level. However, if their response is incorrect, they instead receive a penalty of 0, 2 or 6 marks respectively (Gardner-Medwin, 1998). Thus, more confident responses are rewarded more highly if correct but penalised more severely if incorrect. The particular values used here are calculated so that a person behaving rationally and seeking to maximise their expected score will be motivated to state their true confidence level and not be rewarded for either excessive timidity (underreporting their confidence) or excessive bravado (exaggerating their confidence).

It has been found that university students tend to be poorly calibrated and in general over-estimate their performance, being unaware of their own ignorance (the Dunning–Kruger effect: Ehrlinger et al., 2008), but this seems to improve with practice (Gardner-Medwin & Curtin, 2007). The effect of CA over time should self-correct, because if a CA is well designed a pupil cannot do consistently well by guessing or by systematically over- or under-reporting their confidence level. Studies have found that asking university students to consider how sure they are about an answer prompts them to question why they believe it to be correct, leading to self-checking, self-explanation and higher-level reasoning (Gardner-Medwin & Curtin, 2007). CA has been found to encourage students to focus on understanding rather than just performance and to think more carefully before giving an answer, with students sometimes changing their answer in response to a request for a confidence level (Issroff & Gardner-Medwin, 1998). Little personality or gender effect among university students has been found (Gardner-Medwin and Gahan, 2003).

It seems likely that if CA can be adapted appropriately for school mathematics pupils then some of these benefits might carry across. However, there are possible barriers to implementing CA in schools. Pupils experience anxiety in relation to most kinds of assessment, and this could influence their perceptions of confidence. Negative marking protocols that penalise guessing can be complicated and unfamiliar to school pupils and their teachers. The practice of negative marking can be viewed as punishing pupils for what they do not know, as well as rewarding them for what they do know, and thus could be thought to conflict with a “positive” classroom ethos (Foster, 2007). It might also be feared that CA could weaken pupils’ self-confidence; however, this is desirable when that confidence is misplaced. CA attempts to increase appropriate levels of self-confidence and to support realistic self-awareness in order to facilitate future growth (Dweck, 2000). It is an empirical question to what extent CA might be successfully applied to school mathematics. So, as part of a larger body of work exploring different aspects of mathematical fluency, in this study I seek to answer the research questions:

1. How do school mathematics pupils respond to the use of a CA instrument designed to incentivise the eliciting of truthful confidence ratings? How easy or difficult do they find it to understand? How readily or not do they accept it as a formative assessment tool?
2. What is the relationship between school mathematics pupils’ confidence and competence within the topic of directed (positive and negative) numbers? In particular, how well calibrated are pupils within this topic?

#### **4. Method**

A mixed methods approach was taken so as to take advantage of the complementary strengths of quantitative and qualitative perspectives and mitigate the limitations of each (Ercikan & Roth, 2009). The overall intention was to obtain a more complete understanding of pupils’

responses to the CA instrument than would be provided by either approach alone. A 10-item instrument was designed in order to measure pupils' responses and confidence levels for directed numbers items so that the correlation between facility and confidence could be calculated in order to determine how well calibrated the pupils were. The instrument also contained a space for open comments regarding the process, which would be analysed qualitatively in order to elicit in an open-ended way pupils' perceptions on the convenience, usefulness and acceptability of the CA approach.

#### 4.1 Instrument

The instrument (Figure 2) consisted of 10 items involving calculating with directed numbers (positive and negative numbers and zero). Directed numbers was chosen as the mathematical topic because it is on the school curriculum for a wide range of ages and it was felt that a useful connection might be made between the mathematics and the CA scoring system for the pupils' responses. It was hoped that this would help the pupils to understand the details of the negative marking of the CA approach more easily, since this would be their first encounter with such a system.

Pupils were asked to write down for each item whatever response they thought was correct and to state for each response how confident they were that it was correct. They were asked to give a whole number from 0 to 10 inclusive to indicate how sure they were of each response. It was explained that their total mark would be calculated as the sum of the "how sure" values for the correct responses minus the sum of the "how sure" values for the incorrect responses. It was intended that pupils would regard the confidence scale as linear, since every extra mark that they staked on a response would increase or decrease their total by 1. Although scoring confidence on a 0–10 scale might suggest very fine discernment of confidence levels, a 0–10 scale (rather than, for example, a 1–3 scale, as mentioned above) was intended to simplify the interpretation for the pupils, since the maximum possible score

for the 10 items would be 100, as in more familiar school tests which give a percentage total mark.

#### 4.2 Participants

Fourteen classes of pupils in five secondary schools in three different cities in the UK completed the task. Demographic data on the schools and classes are given in Table 1. The schools involved were a convenience sample and spanned a range of sizes and composition. Teachers were asked to “choose a class with some knowledge of negative numbers, so that the questions on the sheet [Figure 2] might be moderately demanding”. Classes from Year 7 (age 11–12), Year 8 (age 12–13) and Year 9 (age 13–14) were used. In most schools in the UK, mathematics classes are set by attainment, and this was the case for all of the schools in this study apart from school A, which operated a mixed-attainment policy for these year groups. Teachers reported on the attainment of each class, and these were described as “high”, “middle” and “low”, indicating the relevant tertile within the school.

## Directed Numbers

For each of the 10 questions below, write the answer in the “Answer” box.

Each time, also write in the “How Sure?” box a whole number from 0 to 10 to indicate how sure you are that your answer is correct. On this scale 0 is “completely unsure” and 10 is “completely certain”.

	Question	Answer	Leave Blank	How Sure?
1	$5 - 8$			
2	$-2 + 7$			
3	$-3 - 5$			
4	$6 - (-1)$			
5	$4 + (-9)$			
6	$(-3) \times 4$			
7	$5 \times (-2)$			
8	$(-2) \times (-3)$			
9	$\frac{15}{3}$			
10	$\frac{10}{2}$			

Comments

--

Figure 2. Pupil task sheet

School	Location	Size	Gender	Age	Type	Proportion FSM <sup>1</sup>	Proportion minority ethnic <sup>2</sup>	Proportion EAL <sup>3</sup>	Proportion SEND <sup>4</sup>	Most recent Ofsted <sup>5</sup> judgment	Year	Attainment	N
A	London	>	mixed	11-18	comprehensive	>>	>>	>>	>>	good	8	mixed	23
B	London	=	girls	11-18	academy	>	>	>>	<	outstanding	8	low	20
											9	middle	26
C	London	=	mixed	11-16	academy	>	>>	>>	=	outstanding	8	high	22
											8	high	25
											8	middle	17
											8	low	24
											8	low	21
D	East Midlands	<	mixed	11-14	academy	<<	<	<<	=	outstanding	7	high	32
											7	middle	31
											7	low	28
											8	high	25
											8	middle	24
E	West Midlands	<	mixed	11-18	academy	>>	>	>>	>	outstanding	8	high	27

### Notes

<, > and = relate to the national averages in the UK; << and >> indicate well below and well above respectively.

1. proportion of pupils known to be eligible for free school meals

2. proportion of pupils from minority ethnic groups

3. proportion of pupils who are learning English as an additional language

4. proportion of pupils with special educational needs and/or disabilities

5. Ofsted, the Office for Standards in Education, Children's Services and Skills, is a non-ministerial department of the UK government, with the responsibility for inspecting schools and producing reports.

*Table 1. Details of the schools and classes.*

### 4.3 Administration

The instructions given orally to pupils by their teachers were:

Please do this sheet on your own and without using a calculator or any books or number lines. For each question, write down in the “Answer” column whatever answer you think is correct. At the same time, for each of your answers, put in the “How sure?” column how confident you are that it’s correct. Do this by writing a whole number from 0 to 10, where 0 means you have no idea – it’s a complete guess – and 10 means you’re absolutely sure. So if you’re 50% sure, put a 5. Don’t write anything in the “Leave blank” column yet.

Teachers were instructed to ask pupils “Does that make sense?” and answer any questions of clarification about how the sheet should be completed. Then pupils were told: “When you’ve finished, your total mark will be the total of the “How sure?” numbers for the ones you get right minus the total of the “How sure?” numbers for the ones you get wrong.” Again, teachers were instructed to ask pupils “Does that make sense?” and to answer any further questions of clarification.

Teachers were then asked to lead a short discussion intended to help the pupils understand the process by asking “What’s the highest possible mark you could get on this sheet? What’s the lowest?” Pupils were intended to appreciate that if they got every question right and put a 10 for each answer then they could get a total of 100, the highest total possible, whereas if they got every question wrong and put a 10 for each answer then they would get –100.

Pupils were then given as much time as they wanted, individually, without any assistance, to complete the sheet. After this they were asked to change colour of pen (or pencil) and mark their own answers. Teachers were asked to be vigilant at this stage to ensure that pupils did not make any changes to their answers. The pupils were instructed as follows:



You're going to mark the answers in the "Leave blank" column – you don't have to leave it blank any more! Instead of marking with ticks and crosses, you're going to put "+" or "-" into the "Leave Blank" column next to your "How sure?" numbers for each question. Put plus if it's right and minus if it's wrong.

The teacher then read out the correct answers and pupils were reminded how to obtain their total marks:

Now if you look at the sign in front of your "How sure" numbers you can see that they are positive if you got that question right and negative if you got it wrong. So now you add up all of the "How sure?" numbers, taking into account whether they are positive or negative. You can use a calculator if you want.

Finally pupils were given an opportunity to write a comment at the bottom:

Please write something about your total mark. Are you surprised? What do you think of this process where we take account of how sure you are? Please write any comments in the box at the bottom.

They were then thanked for participating in the study, and all of the sheets were returned. Most responses were anonymous and gender was unknown, but some teachers asked the pupils to write their names on the sheets and then the teacher wrote afterwards on each sheet whether the pupil was male or female, in order to enable a gender analysis to take place without at any stage highlighting gender as an issue to the pupils.

Teachers did not report any difficulties in administering the task or any deviations from the instructions given.

## **5. Results**

The data were collated and analysed as detailed below. Facility for each pupil was defined as the number of correct items (out of 10) and confidence for each pupil was defined as the

mean of the confidence ratings for the 10 items. The correlation between facility and confidence was calculated and a  $2 \times 2$  mixed ANOVA was performed to look for a gender effect. Data regarding pupils' attainment sets was not precise enough to enable an analysis. The qualitative data was coded and analysed separately.

### 5.1 Confidence and facility

A total of 345 pupils completed the task. Nine scripts contained missing responses and were excluded, leaving  $N = 336$ . The pupils' facility correlated with their mean confidence with  $r = .546$ ,  $N = 336$ ,  $p < .001$ . This correlation is limited by the variation in each individual pupil's confidence ratings and item-facilities on the different items, which analysis by item reveals (Table 2, Figure 3); yet, despite this, it is large. The internal reliability of the confidence ratings across the 10 items was very good, with a Cronbach's alpha of .900, indicating that within this assessment confidence is a consistent quality, which these items measure coherently. Most items had a fairly high item-facility, and the mean confidence was greater than 5 for all, and greater than 6 for most. Items 4, 8 and 10 offered the greatest challenge. In each case the modal answer was the correct one, and there was an overwhelming correct consensus for items 1, 2, 3, 5, 6, 7 and 9. Items 8 and 10 were bimodal, with the negative of the correct answer being the second most popular choice. For item 4 there were two quite popular incorrect answers (-5 and -7). Pupils were over-confident with items 4 and 8 but under-confident with item 9. It can be seen from Figure 3 that values are close to ceiling for the first three items, so caution is needed when interpreting responses to these.

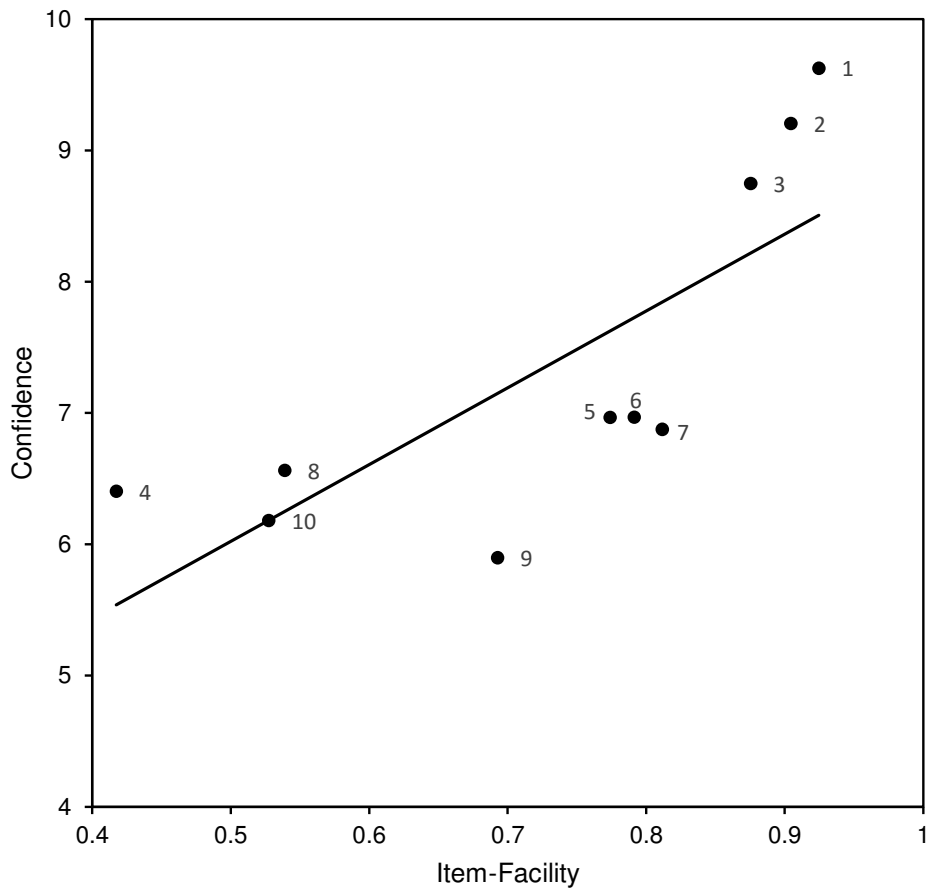


Figure 3. Confidence against item-facility for each item

Item	<i>M</i>	<i>SD</i>
1	.93	.26
2	.90	.29
3	.87	.33
4	.42	.49
5	.78	.42
6	.79	.41
7	.81	.39
8	.54	.50
9	.70	.46
10	.54	.50

Table 2. Mean item-facilities and standard deviations for each item

## 5.2 Gender

Gender was reported for 195 of the pupils (106 female, 89 male). A  $2 \times 2$  mixed ANOVA was performed using within-subjects factor type (facility, confidence) and between-subjects factor gender (male, female). There was a significant interaction between type and gender,

$F(1, 193) = 13.92, p < .001, \eta_p^2 = .067$ , meaning that boys and girls displayed different relationships between facility and confidence. Girls demonstrated lower confidence than their facility would justify whereas boys demonstrated higher confidence than warranted by their facility (see Figure 4). There was no significant main effect of type,  $F(1, 193) = 14.61, p = .167, \eta_p^2 = .010$ , meaning that if we ignore the gender of the pupils then there is no significant difference between facility and confidence; i.e., pupils are well calibrated. There was also no significant main effect of gender,  $F(1, 193) = 2.26, p = .135, \eta_p^2 = .012$ . Post-hoc repeated-measures t-tests confirmed that for the girls ( $N = 106$ ) there was no significant difference between facility ( $M = 6.96, SD = 1.97$ ) and confidence ( $M = 6.63, SD = 1.97$ );  $t(105) = 1.816, p = .072$ . For the boys ( $N = 89$ ), confidence ( $M = 6.80, SD = 1.71$ ) was significantly higher than facility ( $M = 6.08, SD = 2.00$ );  $t(88) = 3.304, p = .001$ . The relative overconfidence of boys is consistent with the findings of Frost, Hyde and Fennema (1994) mentioned earlier.

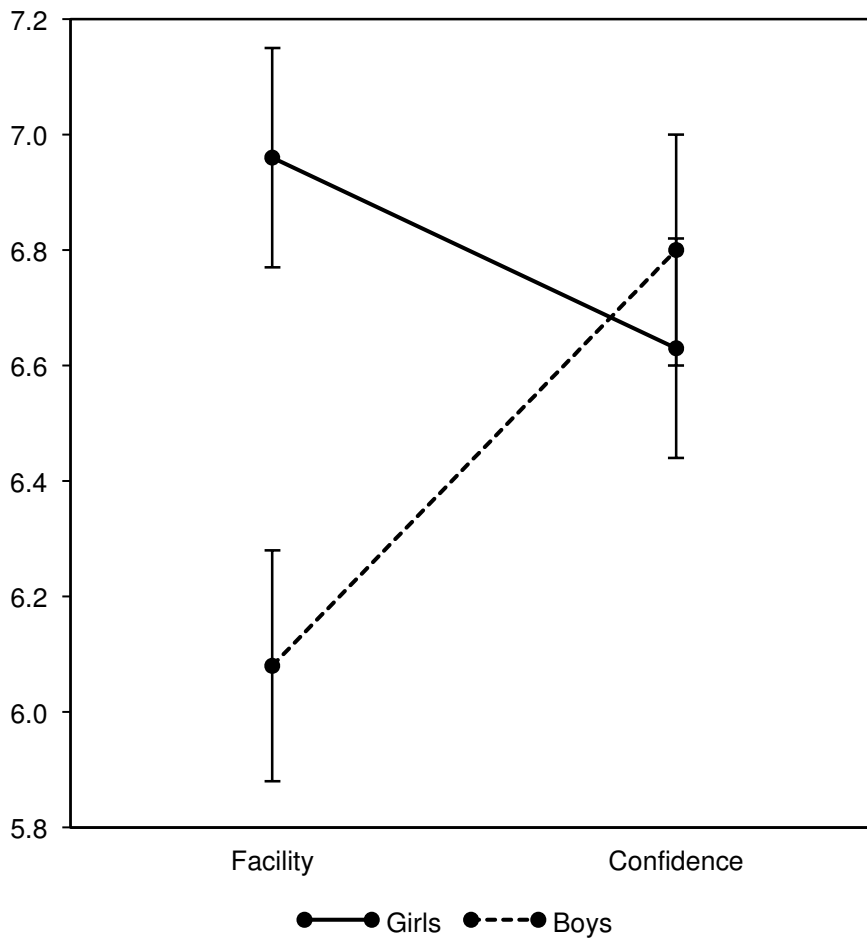


Figure 4. Facility and confidence for girls and boys (error bars show standard errors)

### 5.3 Pupils' views

#### 5.3.1 Overall views

There is evidence that CA has been well received by university medical students (Gardner-Medwin and Gahan, 2003), but what about school mathematics pupils? Pupils' comments were coded according to whether they expressed a positive view of the CA process, a negative view or neither. Many of the comments focused on the pupils' feelings about the topic of directed numbers or did not express a clear opinion either way about the CA approach, and were coded as "neither". Examples of comments coded positively and negatively are:

Positive: “I liked this idea of the worksheet. I like the idea of thinking of how sure you are.” (Pupil 120)

Negative: “I don't like this marking scheme as it is partly based on your confidence in yourself.” (Pupil 217)

Of the 345 pupils there were far more positive responses (106) than negative responses (28). There was no significant difference in the total score (on the –100 to 100 scale) of those expressing positive views ( $M = 44.5$ ,  $N = 106$ ,  $SD = 30.0$ ) and those expressing negative views ( $M = 44.6$ ,  $N = 28$ ,  $SD = 35.7$ );  $t(132) = .015$ ,  $p = .988$ , indicating that pupils were not simply stating a self-interested preference based on their success with the approach. Eight pupils volunteered that they would like to use the approach again; for example, “I really like the how sure ... [a]re you process. I would like to use this more often in class” (Pupil 77). Pupils sometimes indicated that this was because of (rather than despite) its difficulty; for example, “I found this challenging and I think you should do this every lesson” (Pupil 343). Comments were coded in more detail and analysis revealed four overall themes. These themes emerged from analysis of the pupils’ comments and were not decided on beforehand based on the literature.

### 5.3.2 Understanding of the process

Many comments indicated good understanding of how the CA process worked and awareness of the dilemma of wanting to give a high confidence rating if the answer was correct but a low one if it might be incorrect. For example: “I was confident because I wanted points but on ones I wasnt [sic] sure of, I put a low confidence rate [sic]” (Pupil 138). One pupil compared their competence and CA scores directly: “I have got 10/10 but 66/100 due to the lack of confidence in my answer. To get a better answer [sic] I should have felt more confident in my answers and I would have got a better mark” (Pupil 288). One-hundred-and-

eighteen pupils said that they were surprised by the outcome of the CA – many said “very” surprised. Thirty-four pupils commented that their score was low because of their confidence ratings, many indicating that they wished that they had put 10 for the items that they got right.

### 5.3.3 Fairness of the process

Eleven pupils expressed reservations regarding the use of CA, especially if it were to be introduced into high-stakes assessments; for example, “i [sic] don’t like the process of taking ... account of how sure you are because it isn’t fair on people” (Pupil 42). One pupil commented: “I don’t really like the idea of adding up the how sure you were because if someone was really sure of some thing [sic] and got it wrong it would take away 10 so that’s 10 of your mark” (Pupil 182). One pupil questioned the rationale of attempting to measure confidence: “i [sic] don’t think the pross [process] of taking in accant [account] how sure you are is acurat [accurate] because you cant [sic] really put what your [sic] really thinking in numbers” (Pupil 36). Some felt that confidence should be irrelevant to a score in mathematics: “I don’t like the marking scheme because the marks are based on how sure you are and not the work” (Pupil 224). However, others seemed to regard it as fairer than normal marking; for example, “I like the ‘are you sure’ system because even if you are not sure it gives you a chance” (Pupil 175). Some pupils seemed to object to CA mainly because it lowered their score: “I think the process of how sure you are is bad because it makes your score worse” (Pupil 167). But others saw this as good: “I like the process and [it] makes it more challenging to get a high score.” (Pupil 165) The vast majority expressed no concerns regarding fairness.

### 5.3.4 Deeper thinking and increased confidence

Five pupils indicated that they had given more thought to their responses because of the CA process. For example, “I think the ‘How Sure’ column makes you think more” (Pupil 159). Many pupils stressed the importance of confidence; for example, “I was pretty surprised that I got all the answers right, although I am not happy that I am not confident of myself and wasn’t sure of my answers eventhough [sic] they were right” (Pupil 64). Seven pupils stated or implied that the CA process could raise pupils’ confidence. Many stated that they would give higher confidence ratings if they could do a CA again: For example, “I need to be more confident and believe in myself. I only got one wrong. I would be confident next time.” (Pupil 67). Another pupil commented: “I think it’s a Good [sic] idea because it incourges [sic] people to believe in them selfs [sic]” (Pupil 37).

### 5.3.5 Usefulness for formative assessment

Four pupils commented on what they perceived as the prevalence of guessing in the mathematics classroom, implying that CA could address this. For example, “I think that this was a good idea because most of the time kids lie and just guess so this is a good process” (Pupil 4). Two pupils interpreted this from the teacher’s perspective. For example, “I think it’s good to find out how confident people are with their answer because you might guess (and not feel confident) and get it right. This tells the teacher that you’re comfortable, when you’re not” (Pupil 203).

## 6. Discussion

We will now consider the findings under the headings of the research questions given earlier.

6.1 How do school mathematics pupils respond to the use of a CA instrument designed to incentivise the eliciting of truthful confidence ratings? How easy or difficult do they find it to understand? How readily or not do they accept it as a formative assessment tool?



It is clear that pupils understood the CA instrument very readily. The observed large correlation (Cohen, 1992) between facility and confidence implies that pupils were attempting to give true confidence levels, and the clear impression from the comments given was that students were concerned to maximise their score by deploying their confidence ratings appropriately. This means that the confidence responses can potentially be of considerable value to teachers for formative assessment purposes, since they offer a reliable way of probing pupils' confidence at an item-level. However, some caution is needed here because the schools and teachers participating in this study were a convenience sample and may not be representative of schools and teachers more generally.

The vast majority of pupils strongly accepted the CA approach, despite its dramatic differences from usual assessment practices in UK schools. The overwhelmingly positive nature of the comments indicate that on the whole the pupils welcomed its potential benefits to themselves and their teachers. Pupils recognised the importance of assessing their confidence and on the whole felt that the instrument did so fairly. Several pupils suggested that the CA process prompted deeper thinking about their answers and increased their confidence. Further study would be needed to determine whether these perceptions were accurate and to assess teachers' perceptions of the usefulness of the instrument.

It seems likely from this that a CA approach could give teachers reliable data on pupils' genuine confidence levels by disincentivising pupils artificially inflating their expressions of confidence in order to make a positive impression on the teacher. Having more accurate information about the pupils' real confidence levels could enable the teacher to intervene in more effective ways to help move pupils towards better calibration. The CA approach is likely to discourage guessing answers and instead contribute to greater pupil self-awareness, highlighting to pupils, as well as to teachers, when they would benefit from additional support.

6.2 *What is the relationship between school mathematics pupils' confidence and competence within the topic of directed (positive and negative) numbers? In particular, how well calibrated are pupils within this topic?*

Although caution is needed because the schools and teachers used were a convenience sample, the large correlation of .546 between the pupils' facility and their confidence (N = 336,  $p < .001$ ) does indicate that pupils were generally well-calibrated in this topic area. The fact that the correlation, though high, is by no means perfect might be seen positively: teachers should not feel that lowering the difficulty of the questions that they pose is the only way to raise pupils' confidence. Girls demonstrated lower confidence than their facility would justify whereas boys were over-confident, which is consistent with previous findings (Frost, Hyde, & Fennema, 1994). Further study is needed to determine to what extent the calibration and gender effects might be similar across other mathematical topics.

Additionally, facilities were generally high with this instrument and further work will reveal whether similar results are obtained when more difficult items are used and pupils are less sure of their responses. Further work might also reveal any associations with setting practices in schools.

It might be anticipated that with regular use of a CA approach pupils' calibration across various topics would improve over time (Gardner-Medwin & Curtin, 2007), and longitudinal studies would be needed to determine whether this is the case. Many pupils' comments indicated that they would give different confidence ratings if they had the opportunity to attempt another CA, suggesting that they had gained some insight into their reliability in this topic.

## **7. Conclusion**

Mathematics, among all school subjects, presents a uniquely tantalising prospect of certainty. Russell (1956, p. 53) described how in his youth he “wanted certainty in the kind of way in which people want religious faith”, and that he “thought that certainty is more likely to be found in mathematics than elsewhere”. Yet very few school pupils experience mathematics as a place of secure certainties. Mathematics teachers want their pupils to experience the confidence of knowing and understanding mathematics but do not want to engage in assessment practices which encourage pupils to pretend to be confident when they are not. In helping pupils to develop better calibration with regard to procedural competence, CA offers a powerful and easily implemented way to support pupils’ realistic appraisals of their own confidence. Such an approach rewards honest disclosure of confidence and contrasts starkly with “guess-and-hope” strategies that pupils are frequently reported resorting to (Holt, 1990).

It is evident that pupils easily grasped the negative marking aspect of the CA process and showed good calibration. Their comments were overwhelmingly positive about the approach, despite it contrasting strongly with more usual assessment practices in schools. It is encouraging that some pupils felt that CA promoted deeper thinking, increased their confidence and could be useful for formative assessment.

As discussed earlier, the CA approach employed in this study does not attempt to address conceptual confidence. It is possible for a pupil to be very confident that their answer will be marked correct (high procedural confidence) without possessing an underlying sense of confidence in the mathematics behind it. Boaler (2009, p. 121) described a mathematics pupil getting the answer right but not understanding what she was doing: “We just plugged into it. And I think that’s what I really struggled with – I can get the answer, I just don’t understand why”. Such a pupil might show high levels of procedural competence and procedural confidence and, being regarded as a successful pupil, their lack of enjoyment and disinclination to pursue the subject in a post-compulsory phase might appear puzzling to the

teacher. Thus there is a need to expand CA to find ways of eliciting truthful expressions of confidence in concepts, so that the teacher might be assisted in supporting pupils' growth in conceptual confidence too. However, ascertaining and supporting pupils' procedural confidence is an important first step.

## **References**

- Askew, M., Hodgen, J., Hossain, S., & Bretscher, N. (2010). *Values and Variables: Mathematics education in high-performing countries*. London: Nuffield Foundation.
- Bandura, A. (1977). Self-efficacy: toward a unifying theory of behavioral change. *Psychological Review*, 84(2), 191–215.
- Baroody, A.J., Feil, Y., & Johnson, A.R. (2007). An alternative reconceptualization of procedural and conceptual knowledge. *Journal for Research in Mathematics Education*, 38(2), 115–131.
- Black, P., Harrison, C., Lee, C., Marshall, B., & Wiliam, D. (2003). *Assessment for Learning: putting it into practice*. Maidenhead: Open University Press.
- Black, P., & Wiliam, D. (1998). *Inside the Black Box: raising standards through classroom assessment*. London: King's College, London.
- Black, P., & Wiliam, D. (2009). Developing the theory of formative assessment. *Educational Assessment, Evaluation and Accountability*, 21(1), 5–31.
- Boaler, J. (1998). Open and closed mathematics: student experiences and understandings. *Journal for Research in Mathematics Education*, 29(1), 41–62.
- Boaler, J. (2009). *The Elephant in the Classroom: helping children learn and love maths*. London: Souvenir Press Ltd.

- Burton, L. (2004). "Confidence is everything" – perspectives of teachers and students on learning mathematics. *Journal of Mathematics Teacher Education*, 7(4), 357–381.
- Clark, I. (2015). Formative assessment: translating high-level curriculum principles into classroom practice. *Curriculum Journal*, 26(1), 91–114.
- Cohen, J. (1992). A power primer. *Psychological Bulletin*, 112(1), 155.
- Department for Education (DfE) (2013). *Mathematics Programmes of Study: key stage 3, national curriculum in England*. London: DfE.
- Di Martino, P., & Zan, R. (2011). Attitude towards mathematics: a bridge between beliefs and emotions. *ZDM Mathematics Education*, 43(4), 471–482.
- Dweck, C.S. (2000). *Self-Theories: their role in motivation, personality, and development*. Philadelphia: Taylor & Francis.
- Ehrlinger, J., Johnson, K., Banner, M., Dunning, D., & Kruger, J. (2008). Why the unskilled are unaware: further explorations of (absent) self-insight among the incompetent. *Organizational Behavior and Human Decision Processes*, 105(1), 98–121.
- Ercikan, K., & Roth, W.-M. (Eds.). (2009). *Generalizing from Educational Research: beyond qualitative and quantitative polarization*. Abingdon: Routledge.
- Fan, L., & Bokhove, C. (2014). Rethinking the role of algorithms in school mathematics: a conceptual model with focus on cognitive development. *ZDM Mathematics Education*, 46(3), 481–492.
- Fennema, E., & Sherman, J.A. (1976). Fennema-Sherman mathematics attitudes scales: instruments designed to measure attitudes toward the learning of mathematics by females and males. *Journal for Research in Mathematics Education*, 7(5), 324–326.

- Fischhoff, B., Slovic, P., & Lichtenstein, S. (1977). Knowing with certainty: the appropriateness of extreme confidence. *Journal of Experimental Psychology*, 3(4), 552–564.
- Foster, C. (2007, May 11). There's a positive side to negative marking. *Times Educational Supplement*, p. 27.
- Foster, C. (2013). Mathematical études: embedding opportunities for developing procedural fluency within rich mathematical contexts. *International Journal of Mathematical Education in Science and Technology*, 44(5), 765–774.
- Foster, C. (2014). “Can’t you just tell us the rule?” Teaching procedures relationally. In S. Pope (Ed.), *Proceedings of the 8<sup>th</sup> British Congress of Mathematics Education*, April 2014, Vol. 34, No. 2 (pp. 151–158). University of Nottingham.
- Frost, L.A., Hyde, J.S., & Fennema, E. (1994). Gender, mathematics performance, and mathematics-related attitudes and affect: a meta-analytic synthesis. *International Journal of Educational Research*, 21(4), 373–385.
- Galbraith, P., & Haines, C. (1998). Disentangling the nexus: attitudes to mathematics and technology in a computer learning environment. *Educational Studies in Mathematics*, 36(3), 275–290.
- Gardiner, A.D. (2014). Teaching mathematics at secondary level. *The De Morgan Gazette*, 6(1). Retrieved from [http://education.lms.ac.uk/wp-content/uploads/2014/07/DMG\\_6\\_no\\_1\\_2014.pdf](http://education.lms.ac.uk/wp-content/uploads/2014/07/DMG_6_no_1_2014.pdf)
- Gardner-Medwin, A.R. (1995). Confidence assessment in the teaching of basic science. *Research in Learning Technology*, 3(1), 80–85.
- Gardner-Medwin, A.R. (1998). Updating with confidence: do your students know what they don't know? *Health Informatics*, 4, 45–46.

Gardner-Medwin, A.R. (2006). Confidence-based marking: towards deeper learning and better exams. In C. Bryan, & K. Clegg (Eds.), *Innovative assessment in higher education* (pp. 141–149). London: Routledge.

Gardner-Medwin, A.R., & Curtin, N.A. (2007, May). Certainty-based marking (CBM) for reflective learning and proper knowledge assessment. Paper presented at the REAP International Online Conference on Assessment Design for Learner Responsibility [http://www.reap.ac.uk/reap/reap07/Portals/2/CSL/t%20-%20great%20designs%20for%20assessment/raising%20students%20meta-cognition/Certainty\\_based\\_marking\\_for\\_reflective\\_learning\\_and\\_knowledge\\_assessment.pdf](http://www.reap.ac.uk/reap/reap07/Portals/2/CSL/t%20-%20great%20designs%20for%20assessment/raising%20students%20meta-cognition/Certainty_based_marking_for_reflective_learning_and_knowledge_assessment.pdf)

Gardner-Medwin, A.R., & Gahan, M. (2003, July). Formative and summative confidence-based assessment. In J. Christie (Ed.), *Proceedings of the 7th International Computer-Aided Assessment Conference* (pp. 147–155). Loughborough: Loughborough University.

Gattegno, C. (1987). *Science of Education*. New York: Educational Solutions.

Hannula, M.S. (2003). Fictionalising experiences: experiencing through fiction. *For The Learning of Mathematics*, 23(3), 31–37.

Hannula, M.S. (2011). The structure and dynamics of affect in mathematical thinking and learning. In M. Pytlak, E. Swoboda, & T. Rowland (Eds.), *Proceedings of the seventh congress of the European Society for Research in Mathematics Education, CERME 7* (pp. 34–60). Poland: University of Rzesów.

Hannula, M.S., Maijala, H., Pehkonen, E., & Nurmi, A. (2005). Gender comparisons of pupils' self-confidence in mathematics learning. *Nordic Studies in Mathematics Education*, 10(3–4), 29–42.

Hewitt, D. (1996). Mathematical fluency: the nature of practice and the role of subordination. *For the Learning of Mathematics*, 16(2), 28–35.

- Holt, J. (1990). *How Children Fail*. London: Penguin.
- Issroff K., & Gardner-Medwin, A.R. (1998). Evaluation of confidence assessment within optional coursework. In M. Oliver (Ed.), *Innovation in the Evaluation of Learning Technology* (pp. 168–178). London: University of North London.
- Johnston-Wilder, S., & Lee, C. (2010). Mathematical resilience. *Mathematics Teaching*, 218, 38–41.
- Jones, L. (1995). Confidence and mathematics: a gender issue? *Gender and Education*, 7(2), 157–166.
- Kent, G., & Foster, C. (2015). Re-conceptualising conceptual understanding in mathematics. In *Proceedings of the Ninth Congress of European Research in Mathematics Education (CERME9): Thematic Working Group 17* (pp. 98–107). Charles University, Prague.
- Kieran, C. (2013). The false dichotomy in mathematics education between conceptual understanding and procedural skills: an example from algebra. In K.R. Leatham (Ed.), *Vital Directions for Mathematics Education Research* (pp. 153–171). New York: Springer.
- Kyriacou, C. (2005). The impact of daily mathematics lessons in England on pupil confidence and competence in early mathematics: a systematic review. *British Journal of Educational Studies*, 53(2), 168–186.
- Leder, G.C. (1995). Equity inside the mathematics classroom: fact or artifact? In W.G. Secada, E. Fennema, & L.B. Adajian (Eds.), *New directions for equity in mathematics education* (pp. 209–224). Cambridge: Cambridge University Press.
- Leung, F.K. (2014). What can and should we learn from international studies of mathematics achievement? *Mathematics Education Research Journal*, 26(3), 579–605.



- Lim, S.Y., & Chapman, E. (2013). Development of a short form of the attitudes toward mathematics inventory. *Educational Studies in Mathematics*, 82(1), 145–164.
- Morony, S., Kleitman, S., Lee, Y.P. & Stankov, L. (2013). Predicting achievement: confidence vs self-efficacy, anxiety, and self-concept in Confucian and European countries. *International Journal of Educational Research*, 58, 79–96.
- National Council of Teachers of Mathematics (NCTM) (2014). *Procedural Fluency in Mathematics: a position of the National Council of Teachers of Mathematics*. Reston, VA: NCTM.
- Pepin, B., & Roesken-Winter, B. (Eds.). (2014). *From Beliefs to Dynamic Affect Systems in Mathematics Education: exploring a mosaic of relationships and interactions*. London: Springer.
- Pierce, R., & Stacey, K. (2004). A framework for monitoring progress and planning teaching towards the effective use of computer algebra systems. *International Journal of Computers for Mathematical Learning*, 9, 59–93.
- Russell, B. (1956). *Portraits from Memory*. London: George Allen & Unwin.
- Schoendorfer, N., & Emmett, D. (2012). Use of certainty-based marking in a second-year medical student cohort: a pilot study. *Advances in Medical Education and Practice*, 3, 139.
- Stankov, L., Lee, J., Luo, W., & Hogan, D.J. (2012). Confidence: a better predictor of academic achievement than self-efficacy, self-concept and anxiety? *Learning and Individual Differences*, 22(6), 747–758.
- Star, J.R. (2005). Reconceptualizing procedural knowledge. *Journal for Research in Mathematics Education*, 36(5), 404–411.

Star, J.R. (2007). Foregrounding procedural knowledge. *Journal for Research in Mathematics Education*, 38(2), 132–135.

Warwick, P., Shaw, S., & Johnson, M. (2015). Assessment for learning in international contexts: exploring shared and divergent dimensions in teacher values and practices. *Curriculum Journal*, 26(1), 39–69.