International Journal of Educational Research 64 (2014) 49–61

Contents lists available at ScienceDirect

International Journal of Educational Research

journal homepage: www.elsevier.com/locate/ijedures

Calibration of self-evaluations of mathematical ability for students in England aged 13 and 15, and their intentions to study non-compulsory mathematics after age 16[☆]

Richard Sheldrake [*], Tamjid Mujtaba, Michael J. Reiss

Institute of Education, University of London, 20 Bedford Way, London WC1H 0AL, England, United Kingdom

ARTICLE INFO

Article history:
Received 29 March 2013
Received in revised form 21 October 2013
Accepted 28 October 2013
Available online 8 December 2013

Keywords:
Calibration
Confidence
Post-16
Self-beliefs
Self-concept
Subject choice

ABSTRACT

Calibration of mathematics self-evaluations (mathematics task confidence compared against ability) was longitudinally explored through 2490 students from England. Students with accurate task calibration at Year 10 (age 15) reported the highest intentions to study mathematics in Years 12 and 13 (when mathematics is not compulsory), and also generally gave the highest self-reports for further mathematics self-beliefs and attitudes including task-level enjoyment, ease, and interest, and subject-level self-concept. Earlier at Year 8, no differences in intentions were found; over-confident students generally gave the highest self-reports at Year 8, while under-confident students generally gave the lowest self-reports. Gender differences also emerged: girls showed no differences in self-beliefs of ability across calibration groups at Year 10, while accurate boys reported the highest self-beliefs.

1. Introduction

1.1. Self-beliefs

Self-beliefs are influential factors in education. Students' self-concepts (academic subject-specific beliefs of prior ability; Bong & Skaalvik, 2003), for example, have been linked to attainment (Huang, 2011; Marsh, Trautwein, Lüdtke, Köller, & Baumert, 2005) and associated with academic interest (Marsh & Martin, 2011). Self-beliefs have strong influences on students' subject choices, together with past attainment, perceptions of subjects, and numerous other factors (Blenkinsop, McCrone, Wade, & Morris, 2006; Crombie et al., 2005; McCrone, Morris, & Walker, 2005).

Girls have frequently been observed to have lower self-concepts than boys (Fredricks & Eccles, 2002; Marsh, 1989; Rhodes, Roffman, Reddy, & Fredriksen, 2004; Wigfield, Eccles, Mac Iver, Reuman, & Midgley, 1991; Young & Mroczek, 2003). Boys have generally reported higher mathematics self-concepts than girls, even though girls often attain slightly higher (Marsh & Yeung, 1998; Skaalvik & Skaalvik, 2004). Boys have additionally reported higher mathematics self-efficacy (self-beliefs of being able to successfully perform in the future) and intrinsic motivation for mathematics (interest in and enjoyment associated with doing mathematics) compared to girls (Chen, 2003; Nagy, Trautwein, Baumert, Köller, & Garrett, 2006; Pajares & Miller, 1994; Skaalvik & Skaalvik, 2004). Some variations have been found, however, such as when girls

* Corresponding author. Tel.: +44 (0) 20 7612 6000.
E-mail address: rsheldrake@ioe.ac.uk (R. Sheldrake).

0883-0355/$ – see front matter © 2013 The Authors. Published by Elsevier Ltd. All rights reserved.
http://dx.doi.org/10.1016/j.ijer.2013.10.008

valued mathematics higher than boys in Grade 12 after controlling for their perceived mathematics ability (Jacobs, Lanza, Osgood, Eccles, & Wigfield, 2002).

Self-beliefs generally decrease as students increase in age. For primary school students, younger children hold more positive beliefs than older children (Eccles, Wigfield, Harold, & Blumenfeld, 1993; Marsh, 1989). Positive attitudes to science, for example, in primary school students generally decrease with age (Murphy & Beggs, 2003), but peak at age 11 and then subsequently decline, especially in girls (Osborne, Driver, & Simon, 1998). For secondary school students, in general, self-perceptions of ability and subjective task values also decline over time (Jacobs et al., 2002), including self-concepts (De Fraine, Van Damme, & Onghen, 2007). Variations have been observed, however, with mathematics self-concept decreasing from Grades 7 to 9, but then increasing again in Grades 10 and 11 (Marsh, 1989). Gender differences over time have varied across studies, with the gap sometimes increasing (De Fraine et al., 2007), sometimes narrowing (Fredricks & Eccles, 2002), and sometimes remaining constant (Marsh, 1989; Rhodes et al., 2004; Watt, 2004). Gender differences have also emerged in other attitudes, such as girls exhibiting greater declines over time in the intrinsic value they associate with mathematics (Fredricks & Eccles, 2002).

1.2. The accuracy of self-beliefs

Educational research frequently, and often unavoidably, considers self-reported beliefs from participants, introducing a potential discrepancy: self-beliefs may or may not reflect actual abilities. 'Calibration' can be considered as the degree to which beliefs or evaluations (such as self-concept) reflect an actual situation (such as ability evidenced through attainment in examinations); it can measure the overall accuracy of beliefs and the bias or direction of any discrepancy (Hacker, Bol, & Keener, 2008; Lichtenstein, Fischhoff, & Phillips, 1982). Calibration may also be considered an (indirect) indicator of metacognition or the awareness of cognitive processes and related areas, including assessment of personal abilities, knowledge, and task-factors, integral to many theories of learning and self-regulation (Zimmerman, 2000).

Two main theoretical perspectives concerning calibration have arisen (see e.g. Bouffard & Narciss, 2011): self-regulated learning models and social-cognitive motivational theories. Self-regulated learning models (Butler & Winne, 1995) promote accurate calibration as integral to personal well-being and functioning, where self-evaluation has important implications to students' studying approaches and motivation. For example, students may study less if they believe they already master an area, which becomes problematic when this belief is inaccurate (Winne, 1995); accurate reflection and calibration may also allow students to identify and then focus their studies onto their own developing areas of specialism. Social-cognitive theories of behaviour (Bandura, 1989, 1997) provide a contrasting view, where positive calibration biases or over-confidence are a normal state that is not necessarily unproductive or damaging, and which facilitates increased motivation and persistence when difficulties arise, together with providing protection from negative affect. Integrated views are also possible, where the positive or negative effects of calibration biases are contextual, dependant on further factors. In a study of undergraduate students, for example, over-confidence due to motivation to achieve was associated with higher attainment, while over-confidence to defensively self-protect from the negative implications of lower results was associated with lower attainment (Gramzow, Elliot, Asher, & McGregor, 2003).

Inaccurate calibration may not always lead to sustainable learning approaches: continual over-estimation of abilities may eventually result in increased failure; negative affect may occur if results are achieved slower than expected (Carver, 2003); defensively lowering expectations may inadvertently decrease reflective and other personal abilities (Martin, Marsh, & Debus, 2003). Under-confidence can also make students uncritically accept other people's viewpoints or strategies for learning and problem-solving, which may not be personally optimal (Efklides, 2006). Over-confidence may generally have shorter-term benefits at the cost of longer-term well-being (Robins & Beer, 2001).

Primary school students commonly over-evaluate their abilities (Bouffard, Markovits, Vezeau, Boisvert, & Duma, 1998; Bouffard, Vezeau, Roy, & Lengelé, 2011), including their mathematical abilities (Desoete & Roeyers, 2006). While the trend of decreasing self-beliefs over time might then be thought to imply a move from over-evaluation towards accuracy, the situation is more complex and calibration may link with achievement and further factors.

In one study, lower-achieving primary students had less accurately calibrated beliefs of reading ability than higher-achieving students (Bouffard et al., 1998). In a related study, students with over-confident beliefs at Grade 5 (around age 10–11) exhibited similar mathematics performance to those with accurate beliefs, while under-confident students performed worse (Bouffard, Boisvert, & Vezeau, 2003). Students at this age with accurate self-evaluations also had greater increases in satisfaction with their performance compared to students who over-estimated (Narciss, Koerndle, & Dresel, 2011). Boys at this age tended to over-estimate their performance compared to girls, although the girls were not necessarily under-confident (Boekaerts & Rozendaal, 2010). At Grade 6 (around age 11–12), higher-attaining students had higher self-efficacy, self-concept, and more accurately calibrated beliefs (Pajares & Graham, 1999). Chen (2003) applied path analysis to find that calibration accuracy had direct effects on mathematics performance at Grade 7 (around age 12–13), as well as indirect effects mediated through self-efficacy beliefs; no differences in calibration accuracy or bias across genders were found, although boys had higher post-performance self-evaluations than girls (Chen, 2003; Chen & Zimmerman, 2007). Under-confident students at Grade 8 and 9 (around age 13–15) made less mathematics progress during the year and had lower performance approach goals (i.e. whether it was important that peers believe the student to be good at mathematics) compared with accurate and over-evaluating students (Dupeyrat, Escribe, Huet, & Régner, 2011).

Gonida and Leondari (2011) found that girls attained higher mathematics scores than boys at Grade 9 and 10 (around age 14–16), although more girls under-estimated their mathematics performance and more boys over-estimated their performance. The students who over-estimated their performance generally reported significantly more interest in the subject than those who under-estimated for both mathematics and language; those who accurately judged their performance reported more interest in mathematics compared to those who over-estimated, however, and reported the same interest as those who over-estimated for language (Gonida & Leondari, 2011). Higher self-concept and self-concept calibration have also been linked with higher scores for students aged 15 within the Programme for International Student Assessment (PISA) 2000; students over-estimating their self-concept were more likely to have mathematics scores below the country mean, while those under-estimating their self-concept were likely to have mathematics scores above the mean (Chiu & Klassen, 2010).

1.3. The present study: rationale, research questions and hypotheses

Self-beliefs and their accuracy, which relate to subject attainment and even interest, therefore have important implications for secondary school students who are beginning to consider their future specialisms or subject choices.

Key Stage 4 (KS4) in England covers Years 10 and 11, where students are aged from around 14 to 16, and usually ends with students taking General Certificate of Secondary Education (GCSE) or equivalent examinations. Compulsory education in England ended with Year 11 at the time of this investigation; the school leaving age in England subsequently increased to age 17 in 2013 and will increase to age 18 in 2015, although students can undertake part-time studies or work-based learning from age 16 to 17 or 18 rather than necessarily continue at school. Towards the end of KS4, if they plan to continue their studies, students need to select a number of subjects to be studied during Years 12 and 13, usually at Advanced Level General Certificate of Education (A-Level). While the English National Curriculum (which cover ages 5–16) requires that students study mathematics, English, and some form of science at KS4, these subsequently become optional subjects. Increasing the numbers of mathematics and science students and graduates nevertheless remains a priority for England (Royal Society, 2011). Mathematics A-Level entries have only recently recovered following a decline due to curriculum changes in 2002; despite recent improvements, A-Level science subjects have generally seen less (or no) growth in numbers since 1996 compared to other subjects (Department for Education, 2011a). Examination results for England from 2001 to 2012 (JCQ, 2012) also show that on average fewer girls compared to boys sat A-Level mathematics and physics examinations, although girls generally performed equally for mathematics and slightly higher for physics.

Self-beliefs (of ability and expected success) for mathematics, past attainment, and the intrinsic value associated with mathematics have been found to influence mathematics subject choices (Correll, 2001; Watt, 2006). Mathematics subject choices in England have also been influenced by the perceived difficulty of A-Level study and (low) confidence, enjoyment, and perceptions of the personal utility of mathematics (Brown, Brown, & Bibby, 2008; Cann, 2009). Girls were found to be more concerned with being able to cope with A-Level mathematics than boys, for example, while boys were more concerned with the utility of the qualification (QCA Research Faculty, 2007). Girls can see mathematics as being more difficult, which reduces their mathematics self-beliefs and the intrinsic value they associate with mathematics, and subsequently influences their subject choices (Watt, 2006). For equivalent performance, boys typically have higher perceptions of their mathematics ability compared to girls, although mathematics grades have been found to have a significantly larger positive influence on perceived mathematics ability for girls compared to boys (Correll, 2001).

Considering the accuracy of self-beliefs of mathematics ability in relation to subject choices potentially offers a new perspective on these issues. Under-confident students may not select subjects that they might otherwise succeed in and enjoy, for example, while over-confident students may select subjects that they are subsequently unable to continue in. There are further benefits to exploring this area: the relation between calibration and subject choice has received less attention in past research; secondary students (who undertake key examinations and subject choices in Years 10 and 11) have received less attention within calibration research compared to primary, lower-secondary, and undergraduate students; and little calibration research has been undertaken in England, which limits applicability to English policy and practice, and constrains direct comparisons with the existing body of international work.

This study therefore aimed to investigate whether any differences were present in mathematics subject-choice intentions and related factors across broad calibration groups (i.e. students who were under-confident, accurate, or over-confident in their self-evaluated beliefs of their mathematical ability). Over-confident and/or accurate beliefs were hypothesised to be associated with greater inclination to study mathematics further (although potentially for different reasons, depending on the social-cognitive theory and/or the self-regulated learning model). Over-confidence was hypothesised to be associated with higher interest compared to accuracy or under-confidence. Given the varying influence of gender in past studies, the identification and exploration of any gender differences also served as a supplemental aim; girls were expected to under-evaluate their mathematics performance, but further gender hypotheses could not be made from the varying results of previous studies.

2. Method

2.1. Participants

Schools across England were randomly sampled within categorised levels of mathematics attainment and progression (i.e. a matrix of high, average, and low categories for both factors), and those with above-average mathematics attainment and/or progression were over-sampled. Students predicted to attain A* to D grades at GCSE mathematics (approximately the top two-thirds of students) were also targeted within schools, although all others were also able to participate. Such selection allowed greater insight into those students with the likely potential to continue to study post-compulsory mathematics, although influenced wider generalisation. The work was ethically reviewed by the host institution and approved before data collection commenced. Participants completed mathematics-specific questionnaires on two occasions (phases); students were free to decline to participate or to omit responses to any particular question at any time.

From the wider project sample, sets of participants were defined for specific analysis; this exploration focused on a 'younger' longitudinal cohort from England who responded at Year 8 in the first phase and Year 10 in the second (2490 students, 1129 male and 1361 female). Students were followed across any changes of school between the two phases, although this cohort covered the same 89 schools at both phases. The schools were broadly distributed across England: 17% in the East, 2% in the East Midlands, 16% in London, 8% in the North East, 11% in the North West, 25% in the South East, 6% in the South West, 10% in the West Midlands, and 6% in Yorkshire and the Humber.

2.2. Measures

The questionnaires used Likert-type scales to record the participants' self-reported degree of agreement or disagreement with statements covering affective responses to academic subjects, lessons, teachers, and wider views and subject choice intentions (e.g. 'I intend to continue to study maths after my GCSEs'); further items explored personality, general learning orientations, and other influences on learning. Personality measures were only included in the first phase questionnaires. Validated measures were used to inform the questionnaire design, items were reversed where necessary, and the final structures were confirmed by principal component analysis with varimax rotation (for more methodological details, see Reiss et al., 2011).

Measures used here included mathematics academic subject-specific self-concept (or belief of retrospective self-ability; 5 items, e.g. 'I am good at maths', 'I do well in maths tests') and subject-specific intrinsic (7 items, e.g. 'Maths is interesting', 'In maths, it is interesting to find out about the rules and patterns of numbers') and extrinsic (5 items, e.g. 'I think maths is a useful subject', 'I think maths will help me in the job I want to do in the future') motivational beliefs. Further measures included students' emotional responses to mathematics (4 items, e.g. 'When I am doing maths, I am bored', 'When I am doing maths, I get upset'), perceptions of mathematics lessons (7 items, e.g. 'I enjoy my maths lessons', 'In my maths lessons, I have the opportunity to discuss my mathematical ideas'), perceptions of mathematics teachers (11 items, e.g. 'My maths teacher believes that all students can learn maths', 'My maths teacher is good at explaining maths'), advice or pressure to study mathematics (5 items, including the influence of friends, teachers, and family, e.g. 'My teacher thinks that I should continue with maths beyond my GCSEs'), and home support for mathematics achievement (5 items, e.g. 'Someone in my family wants me to be successful at school in maths', 'Someone in my family wants me to talk to them about my maths work'). At the first phase, measures also explored students' competitiveness (8 items, e.g. 'I want to be successful, even if it's at the expense of others', 'A group slows me down'), extroversion (an introversion to extroversion scale, 7 items, e.g. 'I feel comfortable around people', 'I don't like to draw attention to myself'), and internality (an externality to internality locus of control scale, 5 items, e.g. 'When I make plans, I am almost certain to make them work', 'When I get what I want, it's usually because I worked hard for it'). Some individual items (e.g. 'Maths is interesting') did not necessarily distinguish between the taught environment of the academic subject (i.e. school lessons) and the wider domain of mathematics. Combining views of taught lessons with the wider domain or field of mathematics is consistent with earlier studies such as PISA 2003 (OECD, 2005), although other work has experimented with both combining and separating these areas (Marsh et al., 2005). The self-concept measure included a peer-comparison item ('Thinking about your maths lessons, how do you feel you compare with the others in your group'), together with personally-oriented items covering perceptions of ability and prior experiences, in accordance with the commonly established nature of the construct (Marsh, 1992; Wigfield & Eccles, 2000). The reliability of these measures was acceptable; Cronbach's alpha coefficients were generally between .7 and .8 (e.g. first-phase self-concept $\alpha = .821$, intrinsic motivation $\alpha = .790$, extrinsic motivation $\alpha = .702$).

A selection of ability tasks were included at the end of the questionnaires (see Appendix 1 for a sample item), assessing areas perceived to be central to mathematical academic attainment; these included interpreting graphs, explored via the 'racing car' question from PISA 2000 (OECD, 2009), and algebra, explored via questions developed from earlier explorations of mathematical proof (Kuchemann, 2008). Further items recorded the students' confidence in their answers, providing a retrospective self-evaluative assessment of their task ability (e.g. 'How confident are you that your answers to the racing car questions are correct'), together with task-specific measures of enjoyment, ease, and interest. Exploring task-specific ability and self-evaluation in this way differs from considering 'capacity' measures of self-efficacy in advance of completing tasks (e.g. 'How confident are you about solving this question correctly' prior to attempting a solution) or hypothetical or generalised estimates of self-efficacy (e.g. 'How confident are you about solving algebra questions') which have occurred in earlier calibration work. Retrospective self-evaluations of performance, while conceptually distinct from self-efficacy, have been found to be more accurate than predictions (Ackerman & Wolman, 2007), and also cohere with the retrospective nature of self-concept beliefs; strong correlations between self-efficacy and self-evaluation have nevertheless been highlighted (Chen, 2003).

Calibration measures for each student were created following the 'difference score' method from Pajares and Graham (1999), who were informed by and adapted earlier approaches (Keren, 1991; Yates, 1990). Despite the name, the method considers 'mean performance' to be the proportion of correct responses within the tasks, which is not necessarily the same as a 'score' where different item responses are given different values and then summed together. The students' mean confidence and associated mean performance were created on the same scale (1–4; missing responses within the task items were considered as incorrect for this purpose). The calibration 'bias' measure was then formed by subtracting the mean performance from the mean confidence; a positive value then denoted over-confidence, a negative value denoted under-confidence, and a value of zero denoted perfect accuracy in calibration. The bias measure was then converted to a −1 to +1 scale, and a calibration 'accuracy' measure on a 0 to +1 scale was also created by subtracting the absolute (i.e. all values became positive) bias value from 1.

As the calibration bias measure consists of a single continuum of under-confidence through accuracy through to over-confidence, its influence within predictive models is potentially harder to interpret. In order to more easily explore the potential influence of over-confidence or under-confidence, the students were then grouped by each calibration bias measure at each phase: values of −1.00 through −.17 were classified as 'under-confident'; just above −.17 through +.17 were classified as 'accurate'; and just above +.17 through +1.00 were classified as 'over-confident'. These group boundaries allowed a divergence of ±.5 of a 1–4 scale point away from absolute calibration to still be considered as accurate (e.g. a mean performance of 4.00 and confidence of 3.50 provided a bias value of −.1667 which was then classified as 'accurate'; a mean performance of 4.00 and confidence of 3.33 provided a bias value of −.2233 which was then classified as 'under-confident'). This helped ensure that any unavoidable variation introduced through the calculation methods (when averaging across the slightly different number of sub-items within the tasks per phase, for example) was minimised, and the boundary values were still related to the realities of the tasks and question items (which used 4-item confidence scales), while still ensuring a sufficiently informative definition of 'accuracy'.

Methodological decisions made when balancing contextualised issues in practice, including selecting group boundaries, perhaps introduce an unavoidable element of subjectivity, and various approaches could be justified. Different grouping techniques were initially explored, for example, including using more stringent group boundary values of ±.10 (instead of ±.17), and calculating *z*-scores (specific to each of the two phases) from the calibration measures and using ±.5 of a (phase-specific) standard deviation to provide group boundaries. Preliminary analysis highlighted that the main findings reported below were essentially reproduced across these different grouping techniques. The use of phase-specific approaches such as *z*-scores was avoided so as to facilitate direct comparisons between the time points and to enhance any potential generalisation from the results.

SPSS 20 was used for analysis. Initial descriptive statistics and correlations were considered to contextualise the sample, and then group differences were explored. While a stringent level ($p \leq .001$) was preferred for statistical significance, further levels were considered (e.g. $p \leq .01$ and $p \leq .05$) to avoid any potential misrepresentation through omission.

3. Results

3.1. Overall sample calibration accuracy and bias

Tables 1 and 2 provide a descriptive summary of the sample. Paired-sample *t*-tests highlighted that while the overall sample was moderately accurate, absolute task accuracy decreased from Year 8 (.779) to Year 10 (.704; $n = 1949$, SD = .290, $t = 11.403$, 1948df, $p < .001$). For calibration bias, the sample was slightly more over-confident at Year 8 (.040) compared to Year 10 (.015; $n = 1949$, SD = .425, $t = 2.644$, 1948df, $p = .008$). There was low correlation between Year 8 and 10 for calibration bias ($R = .195$, $n = 1949$, $p < .001$) and no correlation for calibration accuracy ($p = .197$). The other main measures moderately correlated between Year 8 and 10 (e.g. self-concept $R = .507$, $n = 2451$, $p < .001$; intrinsic motivation $R = .394$, $n = 2422$, $p < .001$; extrinsic motivation $R = .350$, $n = 2448$, $p < .001$).

3.2. Differences in self-reports across calibration groups

Once categorised into calibration groups (Table 3), group differences across the various measures were explored through Analysis of Variance tests (ANOVAs). Multiple tests were required to ensure equivalence between groups and measures at each phase: when placed into calibration groups resulting from the comparison of task confidence and score at Year 8, differences across the cohort's Year 8 measures were explored; when grouped by calibration at Year 10, differences across Year 10 measures were explored.

Reported intentions to study mathematics further after GCSEs (i.e. at A-Level) did not differ across the task calibration groups at Year 8 ($p = .543$). At Year 10, however, intentions had begun to vary across the groups ($F(2, 1885) = 7.773$, $p < .001$, $\eta^2 = .008$): students in the accurate group were more inclined to study mathematics further (4.33 on a 1–6 scale with 1 denoting strong disagreement and 6 denoting strong agreement) compared with under-confident (4.20) and over-confident (3.99) students; Bonferroni post hoc tests only highlighted a significant difference between the accurate and the over-confident groups ($p < .001$). A further 2×3 (gender × calibration groups) ANOVA test highlighted a gender difference ($p < .001$) for intentions at Year 10 in addition to the calibration group difference, although no interaction occurred ($p = .205$). Further exploration through separate ANOVA tests with the sample split by gender highlighted that the group difference was greater for the boys ($F(2, 844) = 5.524$, $p = .004$, $\eta^2 = .013$; under-confident = 4.47, accurate = 4.67, over-confident = 4.30) compared with the girls ($F(2, 1038) = 4.693$, $p = .009$, $\eta^2 = .009$; under-confident = 4.08, accurate = 3.97, over-confident = 3.69). Interestingly, the pattern of group means differed by gender; Bonferroni post hoc tests only highlighted

Table 1
Descriptive summary of mean responses.

	First phase (Year 8)					Second phase (Year 10)				
				Gender difference					Gender difference	
Measure	All	Boys	Girls	Sig.	*d*	All	Boys	Girls	Sig.	*d*
Mathematics task score (1–4)	2.42	2.45	2.39	N/S	N/S	2.58	2.60	2.56	N/S	N/S
Mathematics task confidence (1–4)	2.72	2.88	2.58	***	.443	2.89	3.11	2.71	***	.494
Mathematics task calibration bias (−1 to +1)	.05	.08	.02	***	.210	.02	.08	−.03	***	.277
Mathematics task calibration accuracy (0 to +1)	.78	.77	.78	N/S	N/S	.70	.73	.68	***	.178
Mathematics task enjoyment	3.63	3.89	3.42	***	.357	3.46	3.66	3.31	***	.239
Mathematics task ease	3.71	4.02	3.46	***	.495	3.90	4.25	3.61	***	.476
Mathematics task interest	3.58	3.76	3.44	***	.234	3.41	3.57	3.28	***	.195
Intention to continue with mathematics after GCSE	4.42	4.54	4.32	***	.154	4.18	4.50	3.91	***	.385
Mathematics academic self-concept	4.14	4.35	3.96	***	.412	4.05	4.34	3.81	***	.535
Mathematics intrinsic motivation	4.17	4.30	4.07	***	.256	3.99	4.13	3.88	***	.269
Mathematics extrinsic motivation	4.87	4.96	4.80	***	.214	4.82	4.94	4.72	***	.266
Emotional response to mathematics	4.04	4.08	4.00	N/S	N/S	3.86	3.95	3.80	***	.146
Perception of mathematics lessons	4.16	4.21	4.12	*	.101	4.11	4.21	4.03	***	.188
Perception of mathematics teachers	4.66	4.62	4.69	*	.088	4.54	4.55	4.53	N/S	N/S
Advice/pressure to study mathematics	4.48	4.55	4.43	*	.100	4.33	4.51	4.18	***	.278
Home support for mathematics achievement	4.78	4.91	4.68	***	.253	4.39	4.56	4.25	***	.365
Competitiveness	4.42	4.28	4.54	***	.375	-	-	-	-	-
Extroversion (introversion/extroversion)	3.96	4.00	3.93	N/S	N/S	-	-	-	-	-
Internality (externality/internality locus of control)	4.25	4.34	4.18	***	.182	-	-	-	-	-

Note: Measures use 1 (strongly disagree) to 6 (strongly agree) scales unless otherwise specified; gender difference *t*-test results: *** represents $p < .001$, **$p < .01$, *$p < .05$, N/S $p > .05$; *d* represents Cohen's *d* measure of effect size.

Table 2
Descriptive summary of correlations.

	Year 8–10 correlations		
Measure	All	Boys	Girls
Mathematics task score (1–4)	.383	.380	.352
Mathematics task confidence (1–4)	.345	.259	.351
Mathematics task calibration bias (−1 to +1)	.195	.197	.174
Mathematics task calibration accuracy (0 to +1)	N/S	.069*	N/S
Mathematics task enjoyment	.377	.345	.383
Mathematics task ease	.352	.306	.320
Mathematics task interest	.325	.264	.365
Intention to continue with mathematics after GCSE	.366	.339	.372
Mathematics academic self-concept	.507	.425	.527
Mathematics intrinsic motivation	.394	.373	.395
Mathematics extrinsic motivation	.350	.328	.353
Emotional response to mathematics	.286	.214	.346
Perception of mathematics lessons	.369	.334	.398
Perception of mathematics teachers	.213	.192	.235
Advice/pressure to study mathematics	.321	.310	.320
Home support for mathematics achievement	.403	.367	.405

Note: All correlations significant at $p < .001$ unless marked N/S (not significant, $p > .05$) or * ($p < .05$).

Table 3
Sample numbers by mathematics task calibration groups.

	First phase (Year 8)					Second phase (Year 10)				
Mathematic task calibration group	Male		Female		Total	Male		Female		Total
Under-confident	186	*(37%)*	313	*(63%)*	499	192	*(30%)*	449	*(70%)*	641
Accurate	445	*(45%)*	550	*(55%)*	995	480	*(53%)*	430	*(47%)*	910
Over-confident	367	*(49%)*	383	*(51%)*	750	285	*(49%)*	300	*(51%)*	585
Total	998	*(44%)*	1246	*(56%)*	2244	957	*(45%)*	1179	*(55%)*	2136

Table 4
First phase (Year 8) measures by first phase (Year 8) mathematics task calibration groups.

	Under-confident (U)			Accurate (A)			Over-confident (O)			Group difference	Bonferroni post hoc difference(s)		
Measure	*M*	SD	*N*	*M*	SD	*N*	*M*	SD	*N*	ANOVA	U – A	U – O	A – O
Mathematics task enjoyment	**3.35**	1.28	497	**3.62**	1.33	984	**3.86**	1.34	700	***	***	***	***
Mathematics task ease	**3.44**	1.13	497	**3.67**	1.18	981	**3.97**	1.14	698	***	***	***	***
Mathematics task interest	**3.37**	1.35	497	**3.58**	1.41	979	**3.76**	1.45	701	***	*	***	*
Intention to continue with mathematics after GCSE	4.38	1.35	486	4.44	1.38	969	4.47	1.36	732				
Mathematics academic self-concept	**4.03**	.97	499	**4.16**	.97	995	**4.24**	.97	750	**	*	***	
Mathematics intrinsic motivation	**4.04**	.84	496	**4.20**	.86	990	**4.24**	.95	745	***	**	***	
Mathematics extrinsic motivation	**4.82**	.74	498	**4.88**	.74	995	**4.93**	.79	749	*		*	
Emotional response to mathematics	**3.95**	.94	498	**4.06**	.97	994	**4.10**	.97	748	*		*	
Perception of mathematics lessons	**4.02**	.92	498	**4.18**	.94	994	**4.27**	.95	748	***	**	***	
Perception of mathematics teachers	**4.53**	.95	499	**4.67**	.87	995	**4.75**	.89	749	***	*	***	
Advice/pressure to study mathematics	**4.36**	1.25	489	**4.54**	1.15	981	**4.52**	1.20	741	*	*	*	
Home support for mathematics achievement	**4.68**	.88	488	**4.78**	.95	977	**4.84**	.96	735	*		*	
Competitiveness	4.43	.71	498	4.46	.70	994	4.41	.70	749				
Extroversion (introversion/extroversion)	**3.91**	.94	498	**3.92**	.92	988	**4.03**	.86	747	*			
Internality (externality/internality locus of control)	**4.05**	.89	498	**4.22**	.90	991	**4.40**	.90	746	***	**	***	***

Note: Measures use 1 (strongly disagree) to 6 (strongly agree) scales; *** represents $p < .001$, **$p < .01$, *$p < .05$; measures with a significant ANOVA difference in group mean scores are also highlighted in bold.

differences between the accurate and the over-confident groups ($p = .003$) for boys, and only between the under-confident and the over-confident groups ($p = .007$) for girls.

Calibration group differences for other measures (summarised in Table 4 for Year 8 and Table 5 for Year 10) were also highlighted. At Year 8, group differences occurred for task enjoyment ($F(2, 2178) = 21.565$, $p < .001$, $\eta^2 = .019$), task ease ($F(2, 2173) = 32.607$, $p < .001$, $\eta^2 = .029$), task interest ($F(2, 2174) = 11.133$, $p < .001$, $\eta^2 = .010$), self-concept ($F(2, 2241) = 6.651$, $p = .001$, $\eta^2 = .006$), intrinsic motivation ($F(2, 2228) = 8.015$, $p < .001$, $\eta^2 = .007$), perceptions of mathematics teachers ($F(2, 2240) = 8.807$, $p < .001$, $\eta^2 = .008$), perceptions of mathematics lessons ($F(2, 2237) = 10.163$, $p < .001$, $\eta^2 = .009$), and internality (locus of control; $F(2, 2232) = 23.569$, $p < .001$, $\eta^2 = .021$). The group differences were less significant for extrinsic motivation ($F(2, 2239) = 3.343$, $p = .036$, $\eta^2 = .003$), advice/pressure to study mathematics ($F(2, 2208) = 4.282$, $p = .014$, $\eta^2 = .004$), emotional responses to mathematics ($F(2, 2237) = 3.813$, $p = .022$, $\eta^2 = .003$), and home support for mathematics achievement ($F(2, 2197) = 4.033$, $p = .018$, $\eta^2 = .004$). Bonferroni post hoc tests generally highlighted significant differences between the under-confident and the accurate groups, and between the under-confident and the over-confident groups; differences only also appeared between the accurate and the over-confident groups for task enjoyment, ease, and interest, and for internality. At Year 10, group differences occurred for task enjoyment ($F(2, 2073) = 8.187$, $p < .001$, $\eta^2 = .008$), task ease ($F(2, 2073) = 19.446$, $p < .001$, $\eta^2 = .018$), self-concept ($F(2, 2133) = 14.300$, $p < .001$, $\eta^2 = .013$), advice/pressure to study mathematics ($F(2, 1955) = 8.132$, $p < .001$, $\eta^2 = .008$), and emotional

Table 5
Second phase (Year 10) measures by second phase (Year 10) mathematics task calibration groups.

	Under-confident (U)			Accurate (A)			Over-confident (O)			Group difference	Bonferroni post hoc difference(s)		
Measure	*M*	SD	*N*	*M*	SD	*N*	*M*	SD	*N*	ANOVA	U – A	U – O	A – O
Mathematics task enjoyment	**3.38**	1.31	631	**3.64**	1.54	883	**3.36**	1.54	562	***	**		**
Mathematics task ease	**3.69**	1.19	633	**4.12**	1.44	883	**3.83**	1.42	560	***	***		***
Mathematics task interest	**3.39**	1.38	627	**3.53**	1.57	877	**3.31**	1.57	559	*			*
Intention to continue with mathematics after GCSE	**4.20**	1.48	556	**4.33**	1.58	822	**3.99**	1.64	510	***			***
Mathematics academic self-concept	**3.92**	.92	641	**4.19**	1.08	910	**4.02**	1.00	585	***	***		**
Mathematics intrinsic motivation	4.00	.83	641	4.06	.93	909	3.95	.94	585				
Mathematics extrinsic motivation	4.81	.73	641	4.87	.82	910	4.80	.83	585				
Emotional response to mathematics	**3.91**	.96	641	**3.95**	1.05	908	**3.76**	1.02	585	**		*	**
Perception of mathematics lessons	4.12	.86	641	4.17	1.00	908	4.06	.98	585				
Perception of mathematics teachers	4.55	.85	641	4.57	.95	906	4.53	.97	583				
Advice/pressure to study mathematics	**4.40**	1.09	584	**4.46**	1.20	832	**4.19**	1.31	542	***		*	***
Home support for mathematics achievement	4.33	.78	625	4.42	.84	888	4.38	.88	566				

Note: measures use 1 (strongly disagree) to 6 (strongly agree) scales; *** represents $p < .001$, **$p < .01$, *$p < .05$; measures with a significant ANOVA difference in group mean scores are also highlighted in bold.

Table 6
Second phase (Year 10) measures by second phase (Year 10) mathematics task calibration groups by gender (selected results only).

	Under-confident (U)			Accurate (A)			Over-confident (O)			Group difference	Bonferroni post hoc difference(s)		
Measure	*M*	SD	*N*	*M*	SD	*N*	*M*	SD	*N*	ANOVA	U – A	U – O	A – O
Mathematics task enjoyment M	**3.43**	1.41	189	**3.86**	1.52	464	**3.55**	1.59	270	**	**		*
Mathematics task enjoyment F	3.36	1.27	442	3.39	1.52	419	3.19	1.48	292				
Mathematics task ease M	**3.88**	1.24	189	**4.51**	1.29	465	**4.14**	1.36	272	***	***		***
Mathematics task ease F	3.61	1.16	444	3.69	1.48	418	3.54	1.41	288				
Intention to continue with mathematics after GCSE M	**4.47**	1.28	168	**4.67**	1.47	429	**4.30**	1.47	250	**			**
Intention to continue with mathematics after GCSE F	**4.08**	1.54	388	**3.97**	1.62	393	**3.69**	1.73	260	**		**	
Mathematics academic self-concept M	**4.09**	.90	192	**4.51**	.95	480	**4.28**	.90	285	***	***		**
Mathematics academic self-concept F	3.85	.92	449	3.84	1.12	430	3.78	1.02	300				
Emotional response to mathematics M	**3.99**	.93	192	**4.09**	1.04	478	**3.81**	1.03	285	**			***
Emotional response to mathematics F	3.88	.97	449	3.80	1.05	430	3.71	1.01	300				
Perception of mathematics lessons M	4.12	.96	192	4.30	.97	478	4.19	.98	285				
Perception of mathematics lessons F	**4.12**	.81	449	**4.03**	1.01	430	**3.95**	.97	300	*		*	
Advice/pressure to study mathematics M	**4.52**	1.08	171	**4.67**	1.11	436	**4.33**	1.21	267	***			***
Advice/pressure to study mathematics F	**4.35**	1.09	413	**4.22**	1.25	396	**4.06**	1.38	275	*		**	

Note: M denotes boys' reports, F denotes girls' reports; measures use 1 (strongly disagree) to 6 (strongly agree) scales; *** represents $p < .001$, **$p < .01$, *$p < .05$; measures with a significant ANOVA difference in group mean scores are also highlighted in bold.

responses to mathematics ($F(2, 2131) = 6.712$, $p = .001$, $\eta^2 = .006$). Group differences were less significant for task interest ($F(2, 2060) = 3.698$, $p = .025$, $\eta^2 = .004$). Bonferroni post hoc tests generally highlighted significant differences between the under-confident and the accurate groups, and between the accurate and the over-confident groups; differences between the under-confident and the over-confident groups only appeared for advice/pressure to study mathematics ($p = .013$) and for emotional responses to mathematics ($p = .031$).

3.3. Gender differences in self-reports across calibration groups

Independent-sample *t*-tests (equal variances were not assumed in all cases for consistency) highlighted that gender differences occurred in most measures, including task calibration bias at both phases. On average, boys showed a higher degree of over-confidence compared to girls at Year 8 (boys = .082, girls = .025; $t = 4.937$, 2142df, $p < .001$, Cohen's $d = .210$) and Year 10 (boys = .077, girls = −.028; $t = 6.403$, 2087df, $p < .001$, $d = .277$). There was no difference in accuracy at Year 8 ($p = .779$) although boys were generally more accurate at Year 10 (boys = .727, girls = .684; $t = 4.040$, 1933df, $p < .001$, $d = .178$). Boys also reported more agreement or positive views than girls for many of the further measures (summarised in Table 1). Gender differences across these (and further) measures have been explored in detail for a wider project sample focusing on physics (Mujtaba & Reiss, 2013a).

Pearson chi-square tests highlighted that gender differences occurred in the composition of the calibration groups at Year 8 ($\chi^2 = 16.538$, 2df, $p < .001$) and Year 10 ($\chi^2 = 84.007$, 2df, $p < .001$). 2 × 3 ANOVA tests then highlighted gender differences in addition to calibration group differences at Year 8 for task enjoyment, ease, and interest (all at $p < .001$), self-concept ($p < .001$), intrinsic motivation ($p < .001$), extrinsic motivation ($p < .001$), perceptions of mathematics teachers ($p = .008$), internality ($p = .001$), and advice/pressure to study mathematics ($p = .040$); no interaction effects were significant. At Year 10, gender differences in addition to calibration group differences occurred for task enjoyment ($p < .001$, where the interaction was also significant at $p = .045$), task ease ($p < .001$, interaction $p = .001$), self-concept ($p < .001$, interaction $p < .001$), emotional responses to mathematics ($p = .001$), and advice/pressure to study mathematics ($p < .001$). While the calibration group difference was not significant for perceptions of mathematics lessons at Year 10, gender ($p < .001$) and the interaction ($p = .022$) were. Table 6 illustrates these statistically-significant (interacting) gender and calibration differences through further analysis of the sample split by gender.

One point of note is that for the girls considered separately, calibration group differences at Year 10 were only significant for intentions to continue mathematics (as described earlier), perceptions of mathematics lessons ($F(2, 1176) = 3.334$, $p = .036$, $\eta^2 = .006$), and advice/pressure to study mathematics ($F(2, 1081) = 4.464$, $p = .012$, $\eta^2 = .008$); post hoc tests highlighted differences between the under-confident and the over-confident groups.

4. Discussion

This work aimed to determine whether mathematics subject-choice intentions and associated factors differed across groups of calibrated self-evaluations of mathematics ability. While no differences in mathematics choice intentions were found at Year 8 (around age 13), the same students at Year 10 (around age 15) illustrated that those who accurately evaluated their task performance had the highest intentions to continue to study mathematics after their GCSEs (i.e. at A-Level in Years 12 and 13 when mathematics is non-compulsory in England). Over-confident students reported the lowest intentions at Year 10, and post hoc tests highlighted the significant difference between the accurate and the over-confident groups.

The work also aimed to explore calibration differences across further factors that potentially influence subject choices. At Year 8, as hypothesised, over-confident students generally reported the highest while under-confident students reported the lowest across the task- and subject-level measures of self-belief and interest; post hoc tests generally highlighted that the under-confident group was distinct, while the other two groups did not generally differ. At Year 10, however, the accurate group generally reported the highest and post hoc tests highlighted that the under-confident and the over-confident groups did not generally differ. These findings offer support to the association of 'costs' with under-confidence (Narciss et al., 2011), at least in younger students; for older students, the findings help support an association between accurate mathematics self-evaluation and higher task-level interest in mathematics (Gonida & Leondari, 2011).

Exploring gender differences was a further aim. Similar to earlier studies of younger students (Boekaerts & Rozendaal, 2010), boys tended to over-evaluate their performance to a greater extent than girls. More girls could generally be placed within the over-confident groups while more boys could generally be placed within the over-confident groups, as also hypothesised, similar to earlier studies with students of approximately the same age (Gonida & Leondari, 2011). Boys also reported higher mathematics self-concept beliefs, intrinsic motivation (i.e. interest in mathematics), and extrinsic motivation (i.e. perceived utility) associated with mathematics, similar to prior research (Nagy et al., 2006; Pajares & Miller, 1994). A new finding was that at Year 10, the measures of mathematics task ease and mathematics self-concept differed for boys across the task calibration groups (with the accurate group reporting the highest beliefs of ease and ability), although no significant differences occurred across the groups for girls. Also at Year 10 when the sample was considered separately for boys and girls, girls' intentions to continue mathematics, perceptions of mathematics lessons, and advice/pressure to study mathematics, followed a different pattern across the groups compared to the boys (although the results were at a lower significance level, so must be interpreted cautiously): across their groups, accurate boys reported the highest, as seen in the full-sample results, while under-confident girls reported the highest. However, the lower statistical significance of these results (and the limitations noted below) suggests that further clarifying research will be necessary. As gender differences associated with calibration have variously been observed (Gonida & Leondari, 2011) or not (Chen, 2003; Chen & Zimmerman, 2007), it appears that this still remains an area for further investigation.

The overall findings suggest that the two contrasting theoretical perspectives of the self-regulated learning model (Butler & Winne, 1995) and the social-cognitive theory of behaviour (Bandura, 1989, 1997) may still be explanatory, but perhaps for different ages and contexts. Biased self-evaluations may be associated with or important for younger students; while mathematics is compulsory in England, an enhanced self-concept may link with attainment (Huang, 2011; Marsh et al., 2005) and academic interest (Marsh & Martin, 2011) as earlier research has highlighted. Such confidence in personal potential, even if this leads technically to being over-confident, may ensure that more effort, persistence, and protection from negative affect occurs (Bandura, 1997). At this stage, however, the results here add support to the view that it is perhaps mathematical under-confidence in students that requires the more attention to help ensure that interest and other motivational beliefs are not also negatively influenced. As students become older the situation may change; higher mathematics self-concept beliefs may be associated more with accurate self-evaluation on the task level, as seen here, rather than over-confidence. The importance of self-regulated learning and accurate self-evaluation may be emphasised or established by the need to select A-Level subjects, for example: students may be prompted to evaluate their strengths, more so than at earlier times, or perhaps receive more tailored feedback and advice. Interestingly, accurate students at Year 8 as well as accurate students at Year 10 reported the highest advice/pressure to study mathematics, which may perhaps proxy for teaching support, including feedback and general comments on their abilities. Self-evaluation skills and accuracy may additionally generally increase with age, a result which has been seen in primary school students at least (Bouffard et al., 1998; Freedman-Doan et al., 2000); accuracy slightly declined in this study from Year 8 to 10, but the tasks may have been too limited to fully determine any changes or consistency at Year 10. The low correlation for task calibration bias between Year 8 and 10 and no correlation for task accuracy suggest that task-level measures of ability or confidence may be inherently more variable as students need to actively engage with material and then self-evaluate each time; in contrast, subject-level ability beliefs such as self-concept may be conceptualised as past-orientated and relatively stable (Bong & Skaalvik, 2003; supported by the stronger correlation seen here between Year 8 and 10 for self-concept). Longitudinal calibration work (e.g. Bouffard et al., 2011) is beginning to suggest that some measure of stability may be present for subject-level calibration biases, although more research will help to clarify the situation.

4.1. Limitations and implications for future research

Further work is required to clarify the associations between calibration, subject-choice, and the included factors through further statistical modelling techniques; the influence of calibration on subject-choices may potentially be smaller than other key factors, such as the perceived utility of a subject and the advice or support provided to students to continue with subjects once they are no longer compulsory, which have been recently highlighted in wider work (Mujtaba & Reiss, 2013b, forthcoming). Regression models (see Appendix 2) highlighted that at Year 10 task calibration bias had an influence on intentions to continue with mathematics after GCSE broadly similar in magnitude to intrinsic motivation, although both factors were notably lower in magnitude than the three most influential factors of the advice or pressure to study mathematics, the extrinsic motivation for mathematics, and mathematics self-concept. Further work is necessary to explore relative influences and magnitudes in detail, especially as task-level measures and subject-level measures may vary in influence when predicting either task-level or subject-level outcomes.

Methodological points are also relevant and indeed are potential limitations of the work. The sample focused on relatively highly-achieving students, who were also from less disadvantaged backgrounds; only around 8% of the sample received free school meals, for example, compared to a national figure of 14–16% during the time of the study (Department for Education, 2011b). Higher-achieving students may have generally more accurate beliefs of their abilities (e.g. Pajares & Graham, 1999), and the relative impact of calibration in the sample may have been influenced accordingly (potentially decreased due to the higher sample uniformity in achievement, for example). The mathematics tasks were also limited in number and scope, and designed to complement a wider exploration into subject choices rather than to undertake a rigorous exploration into calibration itself; more extensive batteries of ability tasks (as applied by e.g. Chen, 2003) would help make task-level calibration results more robust. Any generalisation from the reported findings must be considered cautiously, and further exploration explicitly designed to explore calibration will be necessary to confirm or refute these findings, especially before any recommendations can be made regarding potential interventions to address countering under-confidence in younger (i.e. Year 8) students. With limited resources, interventions aiming to increase the numbers of students studying mathematics may need initially to focus on other areas highlighted by the wider project that encompassed this particular study (e.g. Mujtaba & Reiss, 2013a,b, forthcoming).

Comparing across contextually-specific levels (i.e. using task-level calibration groups to consider differences in subject-level measures) was not ideal (Pajares & Miller, 1995). Effect sizes representing differences across the calibration groups were higher for the task-specific measures of enjoyment, ease, and interest, for example, when compared to the subject-level measures, which perhaps links with calibration also being calculated on the task-level. Although problem-, task-, and subject-specific measures (of self-efficacy at least) highly correlate and were found to equally predict attainment (Bong, 2002), considering the calibration of subject-level self-concept beliefs would help confirm the results presented here. Wider attainment data were unfortunately not available to calibrate the students' self-concept beliefs in this study.

Finally, the calculation of calibration measures can be approached in many ways. As the 'difference score' calibration measures were calculated from the mathematics task performance and task confidence, they could not validly be used to explore calibration differences for those two variables. Additionally, potential variation between the 'difference score' and the 'self-criteria residual' calibration techniques has been highlighted, with a recommendation made to apply both techniques for an enhanced understanding (Gramzow et al., 2003). Various calibration methodologies are indeed possible (Boekaerts & Rozendaal, 2010; Schraw, Kuch, & Gutierrez, 2013), but most explorations (including this one) can only feasibly apply one technique due to practical limitations. Any differences associated with calibration techniques may influence generalisation across studies employing different techniques.

4.2. Conclusions

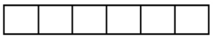
Students with accurate confidence in their mathematics task ability at Year 10 reported the highest intentions to study mathematics into Years 12 and 13 (when mathematics is currently non-compulsory in England), and reported the highest perceptions of enjoyment, ease, and interest in mathematics tasks; under-confident and over-confident students generally reported similarly and slightly lower than accurate students. Earlier at Year 8, however, under-confident students gave the lowest self-reports while over-confident students gave the highest self-reports, including for their perceptions of enjoyment, ease, and interest in mathematics tasks, and for their wider intrinsic and extrinsic motivation for mathematics. Boys tended to over-evaluate their performance to a greater extent than girls; additionally, when considered separately at Year 10, girls showed no differences in self-beliefs of ability (their perceptions of task ease and their wider self-concept beliefs for mathematics) across calibration groups, while accurate boys reported the highest self-concept beliefs and perceptions of task ease.

Acknowledgements

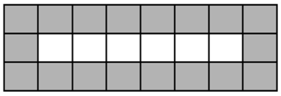
This exploration was undertaken to complement the 'Understanding Participation rates in post-16 Mathematics And Physics' (UPMAP) project, which was supported by the Economic and Social Research Council [grant number RES-179-25-0013]. Please contact the lead author regarding access to the underlying research materials if required.

Appendix 1. A sample mathematics task question

Raj has some white square tiles and some grey square tiles.
They are all the same size.
She makes a row of six white tiles.



She surrounds the white tiles with a single layer of grey tiles.



Then she makes a row of twelve white tiles.

And then she surrounds these white tiles with grey tiles.

a) How many grey tiles does she use to surround a row of 12 white tiles?

Show your working here:

b) How many grey tiles does she need to surround a row of 60 white tiles?

Show your working here:

c) How confident are you that your answer to **part (b)** is correct?

Very confident ☐ Fairly confident ☐ Not confident ☐ Just guessing ☐

d) Write a rule for the number of grey tiles needed to surround a row of n white tiles. You can write a sentence or use algebra.

e) How confident are you that your answer to **part (d)** is correct?

Very confident ☐ Fairly confident ☐ Not confident ☐ Just guessing ☐

Appendix 2. Relative influences on intentions to continue with mathematics after GCSE

	First phase (Year 8)		Second phase (Year 10)	
Measure	β	p	β	p
Mathematics task calibration bias	–.030	.054	**–.052**	**.002**
Mathematics task enjoyment	.022	.469	**.072**	**.019**
Mathematics task ease	**.060**	**.004**	.004	.854
Mathematics task interest	–.008	.765	–.031	.291
Mathematics academic self-concept	**.111**	**<.001**	**.175**	**<.001**
Mathematics intrinsic motivation	.007	.795	**.063**	**.028**
Mathematics extrinsic motivation	**.266**	**<.001**	**.204**	**<.001**
Perception of mathematics lessons	**.229**	**<.001**	.031	.338

Appendix 1 (*Continued*)

Measure	First phase (Year 8) β	First phase (Year 8) p	Second phase (Year 10) β	Second phase (Year 10) p
Emotional response to mathematics lessons	−.004	.826	.024	.280
Perception of mathematics teachers	**−.089**	**<.001**	**−.083**	**<.001**
Advice/pressure to study mathematics	**.339**	**<.001**	**.400**	**<.001**
Home support for mathematics achievement	−.031	.106	.021	.313
Competitiveness	−.009	.577	N/A	N/A
Extroversion (introversion/extroversion)	**−.064**	**<.001**	N/A	N/A
Internality (externality/internality locus of control)	−.023	.183	N/A	N/A

Note: As task calibration bias was formed from task score and confidence, and task calibration accuracy was formed by transforming calibration bias, only calibration bias was included in these models to avoid potential multicollinearity; Year 8 measures were used to predict the reported intentions at Year 8; Year 10 measures were used to predict the reported intentions at Year 10; significant predictors are also highlighted in bold. First phase (Year 8) model: adjusted $R^2 = .524$; model: SE = .931; $F(15, 2026) = 150.981$, $p < .001$. Second phase (Year 10) model: adjusted $R^2 = .520$; model SE = 1.088; $F(12, 1642) = 150.084$, $p < .001$.

References

Ackerman, P., & Wolman, S. (2007). Determinants and validity of self-estimates of abilities and self-concept measures. *Journal of Experimental Psychology: Applied, 13*(2), 57–78 http://dx.doi.org/10.1037/1076-898X.13.2.57

Bandura, A. (1989). Regulation of cognitive processes through perceived self-efficacy. *Developmental Psychology, 25*(5), 729–735 http://dx.doi.org/10.1037/0012-1649.25.5.729

Bandura, A. (1997). *Self-efficacy: The exercise of control*. New York: Freeman.

Blenkinsop, S., McCrone, T., Wade, P., & Morris, M. (2006). *How do young people make choices at 14 and 16? DfES Research Report 773* Nottingham: Department for Education and Skills.

Boekaerts, M., & Rozendaal, J. (2010). Using multiple calibration indices in order to capture the complex picture of what affects students' accuracy of feeling of confidence. *Learning and Instruction, 20*(5), 372–382 http://dx.doi.org/10.1016/j.learninstruc.2009.03.002

Bong, M. (2002). Predictive utility of subject-, task-, and problem-specific self-efficacy judgments for immediate and delayed academic performances. *The Journal of Experimental Education, 70*(2), 133–162 http://dx.doi.org/10.1080/00220970209599503

Bong, M., & Skaalvik, E. (2003). Academic self-concept and self-efficacy: How different are they really? *Educational Psychology Review, 15*(1), 1–40 http://dx.doi.org/10.1023/A:1021302408382

Bouffard, T., Boisvert, M., & Vezeau, C. (2003). The illusion of incompetence and its correlates among elementary school children and their parents. *Learning and Individual Differences, 14*(1), 31–46 http://dx.doi.org/10.1016/j.lindif.2003.07.001

Bouffard, T., Markovits, H., Vezeau, C., Boisvert, M., & Duma, C. (1998). The relation between accuracy of self-perception and cognitive development. *British Journal of Educational Psychology, 68*(3), 321–330 http://dx.doi.org/10.1111/j.2044-8279.1998.tb01294.x

Bouffard, T., & Narciss, S. (2011). Benefits and risks of positive biases in self-evaluation of academic competence: Introduction. *International Journal of Educational Research, 50*(4), 205–208 http://dx.doi.org/10.1016/j.ijer.2011.08.001

Bouffard, T., Vezeau, C., Roy, M., & Lengelé, A. (2011). Stability of biases in self-evaluation and relations to well-being among elementary school children. *International Journal of Educational Research, 50*(4), 221–229 http://dx.doi.org/10.1016/j.ijer.2011.08.003

Brown, M., Brown, P., & Bibby, T. (2008). I would rather die: Reasons given by 16-year-olds for not continuing their study of mathematics. *Research in Mathematics Education, 10*(1), 3–18 http://dx.doi.org/10.1080/14794800801915814

Butler, D., & Winne, P. (1995). Feedback and self-regulated learning: A theoretical synthesis. *Review of Educational Research, 65*(3), 245–281 http://dx.doi.org/10.3102/00346543065003245

Cann, R. (2009). Girls' participation in post-16 mathematics: A view from pupils in Wales. *Gender in Education, 21*(6), 651–669 http://dx.doi.org/10.1080/09540250802680065

Carver, C. (2003). Pleasure as a sign you can attend to something else: Placing positive feelings within a general model of affect. *Cognition and Emotion, 17*(2), 241–261 http://dx.doi.org/10.1080/02699930302294

Chen, P. (2003). Exploring the accuracy and predictability of the self-efficacy beliefs of seventh-grade mathematics students. *Learning and Individual Differences, 14*(1), 79–92 http://dx.doi.org/10.1016/j.lindif.2003.08.003

Chen, P., & Zimmerman, B. (2007). A cross-national comparison study on the accuracy of self-efficacy beliefs of middle-school mathematics students. *The Journal of Experimental Education, 75*(3), 221–244 http://dx.doi.org/10.3200/JEXE.75.3.221-244

Chiu, M. M., & Klassen, R. (2010). Relations of mathematics self-concept and its calibration with mathematics achievement: Cultural differences among fifteen-year-olds in 34 countries. *Learning and Instruction, 20*(1), 2–17 http://dx.doi.org/10.1016/j.learninstruc.2008.11.002

Correll, S. (2001). Gender and the career choice process: The role of biased self-assessments. *American Journal of Sociology, 106*(6), 1691–1730 http://dx.doi.org/10.1086/321299

Crombie, G., Sinclair, N., Silverthorn, N., Byrne, B., DuBois, D., & Trinneer, A. (2005). Predictors of young adolescents' math grades and course enrollment intentions: Gender similarities and differences. *Sex Roles, 52*(5/6), 351–367 http://dx.doi.org/10.1007/s11199-005-2678-1

De Fraine, B., Van Damme, J., & Onghen, P. (2007). A longitudinal analysis of gender differences in academic self-concept and language achievement: A multivariate multilevel latent growth approach. *Contemporary Educational Psychology, 32*(1), 132–150 http://dx.doi.org/10.1016/j.cedpsych.2006.10.005

Department for Education. (2011a). *Maths and science education: The supply of high achievers at A level. Research Report DFE-RR079*. London: Department for Education.

Department for Education. (2011b). *Schools, Pupils, and their Characteristics, January 2011. Statistical First Release SFR 12 2011*. London: Department for Education.

Desoete, A., & Roeyers, H. (2006). Metacognitive macroevaluations in mathematical problem solving. *Learning and Instruction, 16*(1), 12–25 http://dx.doi.org/10.1016/j.learninstruc.2005.12.003

Dupeyrat, C., Escribe, C., Huet, N., & Régner, I. (2011). Positive biases in self-assessment of mathematics competence, achievement goals, and mathematics performance. *International Journal of Educational Research, 50*(4), 241–250 http://dx.doi.org/10.1016/j.ijer.2011.08.005

Eccles, J., Wigfield, A., Harold, R., & Blumenfeld, P. (1993). Age and gender differences in children's self- and task perceptions during elementary school. *Child Development, 64*(3), 830–847 http://dx.doi.org/10.1111/j.1467-8624.1993.tb02946.x

Efklides, A. (2006). Metacognition and affect: What can metacognitive experiences tell us about the learning process? *Educational Research Review, 1*(1), 3–14 http://dx.doi.org/10.1016/j.edurev.2005.11.001

Fredricks, J., & Eccles, J. (2002). Children's competence and value beliefs from childhood through adolescence: Growth trajectories in two male-sex-typed domains. *Developmental Psychology, 38*(4), 519–533 http://dx.doi.org/10.1037/0012-1649.38.4.519

Freedman-Doan, C., Wigfield, A., Eccles, J., Blumenfeld, P., Arbreton, A., & Harold, R. (2000). What am i best at? Grade and gender differences in children's beliefs about ability improvement. *Journal of Applied Developmental Psychology, 21*(4), 379–402 http://dx.doi.org/10.1016/S0193-3973(00)00046-0

Gonida, E., & Leondari, A. (2011). Patterns of motivation among adolescents with biased and accurate self-efficacy beliefs. *International Journal of Educational Research, 50*(4), 209–220 http://dx.doi.org/10.1016/j.ijer.2011.08.002

Gramzow, R., Elliot, A., Asher, E., & McGregor, H. (2003). Self-evaluation bias and academic performance: Some ways and some reasons why. *Journal of Research in Personality, 37*(2), 41–61 http://dx.doi.org/10.1016/S0092-6566(02)00535-4

Hacker, D., Bol, L., & Keener, M. (2008). Metacognition in education: A focus on calibration. In J. Dunlosky & R. Bjork (Eds.), *Handbook of memory and metacognition* (pp. 429–455). Mahwah, New Jersey: Lawrence Erlbaum Associates.

Huang, C. (2011). Self-concept and academic achievement: A meta-analysis of longitudinal relations. *Journal of School Psychology, 49*(5), 505–528 http://dx.doi.org/10.1016/j.jsp.2011.07.001

Jacobs, J., Lanza, S., Osgood, D. W., Eccles, J., & Wigfield, A. (2002). Changes in children's self-competence and values: Gender and domain differences across grades one through twelve. *Child Development, 73*(2), 509–527 http://dx.doi.org/10.1111/1467-8624.00421

JCQ [Joint Council for Qualifications]. (2012). *Examination results*. Retrieved December 16, 2012, from Joint Council for Qualifications: http://www.jcq.org.uk/examination-results.

Keren, G. (1991). Calibration and probability judgements: Conceptual and methodological issues. *Acta Psychologica, 77*(3), 217–273 http://dx.doi.org/10.1016/0001-6918(91)90036-Y

Kuchemann, D. (2008). *Looking for structure: A report of the Proof Materials Project*. London: Unprint Limited.

Lichtenstein, S., Fischhoff, B., & Phillips, L. (1982). Calibration of probabilities: The state of the art to 1980. In D. Kahneman, P. Slovic, & A. Tversky (Eds.), *Judgment under uncertainty: Heuristics and biases* (pp. 306–334). Cambridge: Cambridge University Press.

Marsh, H. (1989). Age and sex effects in multiple dimensions of self-concept: Preadolescence to early childhood. *Journal of Educational Psychology, 81*(3), 417–430 http://dx.doi.org/10.1037/0022-0663.81.3.417

Marsh, H. (1992). Content specificity of relations between academic achievement and academic self-concept. *Journal of Educational Psychology, 84*(1), 35–42 http://dx.doi.org/10.1037/0022-0663.84.1.35

Marsh, H., & Martin, A. (2011). Academic self-concept and academic achievement: Relations and causal ordering. *British Journal of Educational Psychology, 81*(1), 59–77 http://dx.doi.org/10.1348/000709910X503501

Marsh, H., & Yeung, A. S. (1998). Longitudinal structural equation models of academic self-concept and achievement: Gender differences in the development of math and english constructs. *American Educational Research Journal, 35*(4), 705–738 http://dx.doi.org/10.3102/00028312035004705

Marsh, H., Trautwein, U., Lüdtke, O., Köller, O., & Baumert, J. (2005). Academic self-concept, interest, grades, and standardized test scores: Reciprocal effects models of causal ordering. *Child Development, 76*(2), 397–416 http://dx.doi.org/10.1111/j.1467-8624.2005.00853.x

Martin, A., Marsh, H., & Debus, R. (2003). Self-handicapping and defensive pessimism: A model of self-protection from a longitudinal perspective. *Contemporary Educational Psychology, 28*(1), 1–36 http://dx.doi.org/10.1016/S0361-476X(02)00008-5

McCrone, T., Morris, M., & Walker, M. (2005). *Student choices at Key Stage 3 – Literature review. DfES Research Report RW68*. London: Department for Education and Skills.

Mujtaba, T., & Reiss, M. J. (2013a). What sort of girl wants to study physics after the age of 16? Findings from a large-scale UK Survey. *International Journal of Science Education, 35*(17), 2979–2998 http://dx.doi.org/10.1080/09500693.2012.681076

Mujtaba, T., & Reiss, M. J. (2013b). A survey of psychological, motivational, family and perceptions of physics education factors that explain 15-year-old students' aspirations to study physics in post-compulsory English schools. *International Journal of Science and Mathematics Education* http://dx.doi.org/10.1007/s10763-013-9404-1

Mujtaba, T., Reiss, M. J., Rodd, M., & Simon, S. (forthcoming). Methodological issues in mathematics education research when exploring issues around participation and engagement. In J. Middleton & J. Cai (Eds.), *Design, Results, and Implications of Large-Scale Studies in Mathematics Education*. Dordrecht: Springer.

Murphy, C., & Beggs, J. (2003). Children's perceptions of school science. *School Science Review, 84*(308), 109–116.

Nagy, G., Trautwein, U., Baumert, J., Köller, O., & Garrett, J. (2006). Gender and course selection in upper secondary education: Effects of academic self-concept and intrinsic value. *Educational Research and Evaluation, 12*(4), 323–345 http://dx.doi.org/10.1080/13803610600765687

Narciss, S., Koerndle, H., & Dresel, M. (2011). Self-evaluation accuracy and satisfaction with performance: Are there affective costs or benefits of positive self-evaluation bias? *International Journal of Educational Research, 50*(4), 230–240 http://dx.doi.org/10.1016/j.ijer.2011.08.004

OECD [Organisation for Economic Co-operation and Development]. (2005). *PISA 2003 Technical Report*. OECD Publishing.

OECD [Organisation for Economic Co-operation and Development]. (2009). *PISA take the test: Sample questions from OECD's PISA assessments*. OECD Publishing.

Osborne, J., Driver, R., & Simon, S. (1998). Attitudes towards science: Issues and concerns. *School Science Review, 79*(288), 27–33.

Pajares, F., & Graham, L. (1999). Self-efficacy, motivation constructs, and mathematics performance of entering middle school students. *Contemporary Educational Psychology, 24*, 124–139 http://dx.doi.org/10.1006/ceps.1998.0991

Pajares, F., & Miller, M. D. (1994). Role of self-efficacy and self-concept beliefs in mathematical problem solving: A path analysis. *Journal of Educational Psychology, 86*(2), 193–203 http://dx.doi.org/10.1037//0022-0663.86.2.193

Pajares, F., & Miller, M. D. (1995). Mathematics self-efficacy and mathematics performances: The need for specificity of assessment. *Journal of Counseling Psychology, 42*(2), 190–198 http://dx.doi.org/10.1037/0022-0167.42.2.190

QCA [Qualifications and Curriculum Authority] Research Faculty. (2007). *Evaluation of participation in GCE mathematics. Final Report. QCA/07/3388*. London: Qualifications and Curriculum Authority.

Reiss, M. J., Hoyles, C., Mujtaba, T., Riazi-Farzad, B., Rodd, M., Simon, S., et al. (2011). Understanding participation rates in post-16 mathematics and physics: Conceptualising and operationalising the UPMAP Project. *International Journal of Science and Mathematics Education, 9*(2), 273–302 http://dx.doi.org/10.1007/s10763-011-9286-z

Rhodes, J., Roffman, J., Reddy, R., & Fredriksen, K. (2004). Changes in self-esteem during the middle school years: A latent growth curve study of individual and contextual influences. *Journal of School Psychology, 42*(3), 243–261 http://dx.doi.org/10.1016/j.jsp.2004.04.001

Robins, R., & Beer, J. (2001). Positive illusions about the self: Short-term benefits and long-term costs. *Journal of Personality and Social Psychology, 80*(2), 340–352 http://dx.doi.org/10.1037/0022-3514.80.2.340

Royal Society. (2011). *Preparing for the transfer from school and college science and mathematics education to UK STEM higher education: A 'state of the nation' report*. London: Royal Society.

Schraw, G., Kuch, F., & Gutierrez, A. (2013). Measure for measure: Calibrating ten commonly used calibration scores. *Learning and Instruction, 24*, 48–57 http://dx.doi.org/10.1016/j.learninstruc.2012.08.007

Skaalvik, S., & Skaalvik, E. (2004). Gender differences in math and verbal self-concept, performance expectations, and motivation. *Sex Roles, 50*(3/4), 241–252 http://dx.doi.org/10.1023/B:SERS.0000015555.40976.e6

Watt, H. (2004). Development of adolescents' self-perceptions, values, and task perceptions according to gender and domain in 7th- through 11th-grade Australian students. *Child Development, 75*(5), 1556–1574 http://dx.doi.org/10.1111/j.1467-8624.2004.00757.x

Watt, H. (2006). The role of motivation in gendered educational and occupational trajectories related to maths. *Educational Research and Evaluation, 12*(4), 305–322 http://dx.doi.org/10.1080/13803610600765562

Wigfield, A., & Eccles, J. (2000). Expectancy–value theory of achievement motivation. *Contemporary Educational Psychology, 25*(1), 68–81 http://dx.doi.org/10.1006/ceps.1999.1015

Wigfield, A., Eccles, J., Mac Iver, D., Reuman, D., & Midgley, C. (1991). Transitions during early adolescence: Changes in children's domain-specific self-perceptions and general self-esteem across the transition to junior high school. *Developmental Psychology, 27*(4), 552–565 http://dx.doi.org/10.1037/0012-1649.27.4.552

Winne, P. (1995). Inherent details in self-regulated learning. *Educational Psychologist, 30*(4), 173–187 http://dx.doi.org/10.1207/s15326985ep3004_2

Yates, F. (1990). *Judgment and decision making*. Englewood Cliffs, New Jersey: Prentice Hall.

Young, J., & Mroczek, D. (2003). Predicting intraindividual self-concept trajectories during adolescence. *Journal of Adolescence, 26*(5), 586–600 http://dx.doi.org/10.1016/S0140-1971(03)00058-7

Zimmerman, B. (2000). Attaining self-regulation: A social–cognitive perspective. In M. Boekaerts, P. Pintrich, & M. Zeidner (Eds.), *Handbook of self-regulation* (pp. 13–39). Burlington, MA: Elsevier Academic Press.