International Journal of Educational Research 116 (2022) 102085

Contents lists available at ScienceDirect

International Journal of Educational Research

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

Implications of under-confidence and over-confidence in mathematics at secondary school

Richard Sheldrake *, Tamjid Mujtaba, Michael J. Reiss

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

ARTICLE INFO

Keywords:
Calibration
Confidence
Self-concept
Mathematics
Germany

ABSTRACT

Confidence is theorised to be motivational and beneficial within education, although it remains unclear how calibration bias (the extent of under-confidence to over-confidence) might arise and what the implications may be. In order to gain new insights, a longitudinal sample of 3203 secondary school students in Germany was considered at Grade 5 and Grade 9. Predictive modelling explored what factors predicted calibration bias, and whether/how calibration bias predicted other outcomes. The results offered many new insights including that, at Grade 9, calibration bias (i.e. higher over-confidence) negatively predicted mathematics grades but positively predicted mathematics self-concept (subject-level confidence), school self-concept (general educational confidence), and self-esteem (an indicator of well-being), accounting for students' background characteristics and an array of other predictors.

1. Introduction

Students' confidence in their own abilities can be motivational and beneficial within their education (Bandura, 1997). In Germany and other countries, students' confidence and other attitudes related to mathematics have associated with important educational outcomes, including their future attainment and their studying intentions and choices (Guo et al., 2016; Nagy et al., 2006; Regan & DeWitt, 2015; Sheldrake et al., 2015). Mathematical confidence and numerical skills are particularly relevant to everyday life, such as within employment and when managing personal finances (OECD, 2015c), and especially for careers in science-related areas (Science, Technology, Engineering, Mathematics (STEM) / Mathematik, Informatik, Naturwissenschaft, Technik (MINT)) (Anger et al., 2022; Autorengruppe Bildungsberichterstattung, 2022). Enhancing students' mathematics performance, and increasing access to careers in science-related areas including mathematics, are priorities for Germany and many other countries (Acatech, 2022; Anger et al., 2022; EACEA, 2015, 2022).

Students' confidence may not always correspond to their actual performance, however: students can exhibit various degrees of under-confidence or over-confidence, through expressing lower or higher confidence than expected given their performance (Bouffard & Narciss, 2011; Freund & Kasten, 2012; Zell & Krizan, 2014). Theoretical perspectives propose that under-confidence may be limiting to motivation (Bandura, 1997), and research has indeed revealed that under-confident students have held lower attitudes towards mathematics than other students, despite having similar or higher performance (Gonida & Leondari, 2011; Narciss et al., 2011; Sheldrake et al., 2014). Earlier studies have tended to explore whether similarities or differences can be observed across groups of under-confident or over-confident students, so it remains less clear how under-confidence or over-confidence may arise and what their

* Corresponding author.
E-mail address: r.sheldrake@ucl.ac.uk (R. Sheldrake).

https://doi.org/10.1016/j.ijer.2022.102085
Received 10 June 2022; Received in revised form 1 September 2022; Accepted 7 October 2022

wider implications may be. Applying predictive modelling to consider the extent of under-confidence to over-confidence may offer new or complementary insights to help extend understanding. Greater clarity could have important educational implications, such as helping inform whether and/or how educators could intervene.

The presented research aimed to provide new knowledge and insight by considering a longitudinal sample of secondary school students in Germany from Grade 5 to Grade 9 (10/11 years old to 14/15 years old), where Grade 9 is the last year of secondary education before upper-secondary education. The study used predictive modelling to consider whether/how aspects of education and life associated with extents of under-confidence/over-confidence at Grade 5 and Grade 9, which (essentially) considered what might link with (or potentially lead to) under-confidence/over-confidence at different times. Similarly, the modelling also explored whether/how extents of under-confidence/over-confidence predicted other outcomes at Grade 5 and Grade 9, including school self-concept (an indicator of confidence in education in general), self-esteem (an indicator of well-being), and life satisfaction (another indicator of well-being).

1.1. Students' confidence

Confidence has been conceptualised and measured in various ways within educational and motivational research. Someone's current confidence (often conceptualised as 'self-concept beliefs') broadly considers their subjective interpretations of their historic and current performance, often through evaluations of whether someone thinks that they are 'doing well' or are 'good' at an academic subject (Bong & Skaalvik, 2003; Shavelson et al., 1976). Nevertheless, confidence conceptualised as self-concept beliefs does not consider someone's wider sense of self or identity. Self-concept beliefs are also different to self-esteem, which reflects someone's subjective sense of satisfaction and worth in life, essentially providing an indicator of subjective well-being (Bong & Skaalvik, 2003; Rosenberg, 1965; Shavelson et al., 1976). Someone's confidence in their future capacities (often conceptualised as 'self-efficacy beliefs') considers their evaluative beliefs about specific events or contexts, such as their confidence in being able to gain a particular examination grade or in being able to successfully undertake a particular type of task (Bandura, 1977; Bong & Skaalvik, 2003). Confidence can also be measured in various other ways, often involving simplified or contextualised self-evaluations that may be orientated to consider the past (such as 'How confident are you that your answer was correct?' or 'How many questions do you think you answered correctly on this test?') or orientated to consider the future (such as 'How many questions do you expect to answer correctly on this test?').

Someone's confidence, whether considered through self-concept, self-efficacy, or other conceptualisations or measures, may be influenced by many aspects of education and wider life, such as receiving praise (Bong & Skaalvik, 2003; Efklides, 2011). Self-efficacy beliefs have often been found to be influenced by mastery experiences (such as actual performance), vicarious experiences (seeing other people succeed), social persuasions (such as praise), and physiological states (such as anxiety) (Bandura, 1997; Bong & Skaalvik, 2003; Usher and Pajares, 2008b). Self-concept beliefs have often been inferred to be influenced by implicit comparisons against peers and comparisons across subjects (Marsh et al., 2014; Möller & Marsh, 2013; Möller et al., 2011). It is possible that someone's subjective and varying interpretations, evaluations, and various wider influences (and/or other aspects of life) may potentially foster or facilitate 'confidence biases' towards under-confidence or over-confidence (considered as lower or higher confidence than would be expected given someone's performance). Nevertheless, it remains less clear how/why confidence biases arise: more empirical research has focused on predictive modelling of confidence beliefs (whether self-concept, self-efficacy, or other indicators) and less research has focused on predictive modelling of confidence bias (the extent of under-confidence to over-confidence).

Within social-cognitive theory, someone holding higher confidence in their personal abilities is proposed to be motivational and may facilitate them to surpass their normal performance, while low confidence may be limiting and may mean that some actions are not even attempted (Bandura, 1997). Across various studies, higher confidence has indeed associated with motivational approaches such as aiming to learn and master academic work (Jiang et al., 2014; Phillips & Gully, 1997) and persistence (Multon et al., 1991; Skaalvik et al., 2015). Students' confidence has also associated with important educational outcomes, including their subsequent attainment and their studying intentions and choices (Guo et al., 2016; Jansen et al., 2015; Nagy et al., 2006; Taskinen et al., 2013; Trautwein et al., 2012). For example, higher confidence has associated with slightly higher subsequent attainment, even while accounting for prior attainment (Huang, 2011; Möller et al., 2011; Seaton et al., 2014; Talsma et al., 2018). In general, students' confidence and also students' performance at school have linked with indicators of their well-being such as self-esteem and life satisfaction (Gustafsson et al., 2010; Suldo et al., 2006).

Within self-regulation theories of studying and learning, an accurate awareness of personal abilities has been proposed to be beneficial (Boekaerts, 1999; Butler & Winne, 1995; Winne, 1995; Zimmerman, 2000; Zimmerman & Moylan, 2009). For example, reflecting on whether material has been understood and developing an accurate evaluation of personal abilities and understanding may help someone to consider whether they need to revise more in order to gain a particular outcome, such as a specific grade in an important examination. If someone is already confident in their abilities then it is possible that they may feel less need to study or revise further, which may be problematic if they are over-confident (Winne, 1995). Higher confidence has indeed associated with students' self-regulation of their own learning (Usher and Pajares, 2008a; Zimmerman & Schunk, 2011), although it remains less clear (via empirical research) whether over-confidence might help or hinder self-regulation and other outcomes.

Overall, existing research and theoretical perspectives emphasise the importance of students' confidence, although it remains unclear how/why under-confidence/over-confidence may arise and whether these might be beneficial. Nevertheless, social-cognitive theory and self-regulation theory are not necessarily mutually exclusive; specifically, both perspectives suggest potential benefits arising from higher confidence (assuming that it is accurately-evaluated). It is also possible that different theoretical perspectives may help explain different circumstances.

1.2. Under-confidence and over-confidence

Under-confidence and over-confidence have been called 'confidence biases', 'calibration biases', 'self-evaluation biases', and other similar terms (Bouffard & Narciss, 2011; Lichtenstein & Fischhoff, 1977; Stone, 2000). Research has explored the potential implications of confidence biases in order to consider which circumstances might be beneficial and which might be detrimental (and hence to reveal particular cases that may require support within teaching and learning). Accordingly, various studies have defined 'under-confident', 'accurate', and 'over-confident' groups of students, and have considered any differences in their reported attitudes and beliefs. For example, for primary school students in Canada (across Grade 3 and Grade 5; around age 9 and 11 years old), under-confident students reported lower intrinsic motivation (inherent interest and enjoyment) for mathematics, pride in their results, and attitude to effort (i.e. being less likely to apply effort as a means to succeed), and gained lower attainment in mathematics, compared to accurate and over-confident students (Bouffard et al., 2003). For students at the start of secondary school in Germany (Grade 5, age 10/11 years old), under-confident students exhibited lower satisfaction with their mathematics performance than other students, despite gaining the same performance (Narciss et al., 2011). For secondary school students in Greece (across Grade 9 and Grade 10, age 15-16 years old), accurately-evaluating students reported the highest interest in mathematics, then over-confident students, and then under-confident students (Gonida & Leondari, 2011). For secondary school students in England in Year 8 (age 13 years old), under-confident students reported lower attitudes towards mathematics than accurate and over-confident students, including for their interest in mathematics; however, at Year 10 (age 15 years old), fewer differences were apparent but over-confident students reported the lowest intentions to study mathematics further (Sheldrake et al., 2014). Considered more generally, research studies of students of different ages within different countries (often focusing on university students) have often found that higher performing students have been more accurate in their confidence while lower performing students have often exhibited greater over-confidence (for example: García et al., 2016; Kruger & Dunning, 1999; Lichtenstein & Fischhoff, 1977).

It remains less clear how prevalent under-confidence or over-confidence might be. Considered generally, students of varying ages from primary school through to university have often (but not universally) exhibited some degree of over-confidence, when considered on average across samples (for example: Bouffard et al., 1998; Bouffard et al., 2011; Desoete & Roeyers, 2006; Gramzow et al., 2003; Kruger & Dunning, 1999; Winne & Jamieson-Noel, 2002). Some studies have started to gain greater insight through identifying naturally-emergent clusters of students with distinct magnitudes of confidence accuracy/bias, while also considering changes over time. For example, a study of Canadian students from Grades 3/4 to 7/8 (age around 9 to 14 years old) revealed that most students had relatively stable accuracy/bias over time (when comparing their general perceptions of ability against intelligence test scores): most showed a stable but moderate over-estimating bias (75% of the sample), while others showed a stable but highly over-estimating bias (15%); a minority (6%) showed an under-estimating bias that became even more under-estimating over time, while another minority (4%) showed a highly under-estimating bias that changed to a moderately under-estimating bias over time (Bouffard et al., 2011). The students that exhibited the stable but highly over-estimating bias reported higher self-esteem than all others, and performed higher than others in mathematics and languages (except for performing similarly in mathematics to those with a moderate over-estimating bias) (Bouffard et al., 2011). Another study of Canadian students from Grades 4/5 to 9/10 (age around 11 to 16 years old) also highlighted relatively stable accuracy/bias over time (considering students' generalised beliefs of being good at school compared against standardised test scores), with students being either over-confident (28% of the sample), accurate (55%), or under-confident (17%) (Bonneville-Roussy et al., 2017). Specifically considering mathematics, one study of primary school students in Finland (Grade 1 to 2, age 6/7 to 7/8 years old) also revealed naturally-emergent clusters: half of the students (49%) were over-confident with high mathematics self-concept beliefs but low mathematics attainment, a third (33%) were accurate with high beliefs and high attainment, and the remainder (18%) were accurate with low beliefs and low attainment (Rytkönen et al., 2007). Cluster membership was relatively stable over time, although the proportion of students in the over-confident group slightly decreased while the proportion of students in the accurate clusters slightly increased (Rytkönen et al., 2007). Another study considering secondary school students in Spain (Grade 10 to 11, age 15 to 16 years old) revealed clusters of students with accurately-high mathematics self-concept beliefs (10% of the students), accurately-low beliefs (21%), over-confident beliefs (18%), and under-confident beliefs (20%), where cluster membership remained relatively constant over time (Sáinz & Upadyaya, 2016).

Overall, many studies of primary and secondary school students have considered groups or clusters of students. It is reasonable to infer that any differences seen across groups or clusters may associate with under-confidence/over-confidence. Expressed another way, the differences could be inferred to be potential antecedents or potential implications of under-confidence/over-confidence (although causal associations can only be inferred). For example, given that under-confident students have exhibited lower satisfaction with mathematics performance than other students (Narciss et al., 2011), it is possible to infer that lower satisfaction with mathematics performance is either an antecedent and/or an outcome of under-confidence.

1.3. Research aims

Enhancing students' mathematics performance and motivation is a priority for many countries across Europe including Germany (EACEA, 2015, 2022; KMK, 2022a, 2022b). Across Germany inequalities occur within education, including where students with less advantaged circumstances often have lower performance compared to others (Autorengruppe Bildungsberichterstattung, 2020, 2022). Specifically in mathematics, those with more advantaged circumstances have tended to increase their mathematics competencies at a greater rate than others (Autorengruppe Bildungsberichterstattung, 2022). Girls and boys can also have varying progression rates: girls often start primary school with lower mathematical competencies than boys but gain at a greater rate, such that boys and girls then have similar competences at the end of primary school; however, girls often have lower mathematical competencies than boys across secondary school where boys gain at a greater rate (Autorengruppe Bildungsberichterstattung, 2022). Those with higher mathematics performance often enter science-related careers (Autorengruppe Bildungsberichterstattung, 2022).

Considering confidence in more detail may offer insights to help support students within education, including within the context of existing inequalities. For example, girls have often conveyed lower confidence in mathematics than boys (OECD, 2015b), which may sometimes involve under-confidence (Gonida & Leondari, 2011; Sheldrake et al., 2014). Generally, higher confidence has been theorised to be motivational (Bandura, 1997), and has indeed associated with various benefits within education (for example: Guo et al., 2016; Jiang et al., 2014; Nagy et al., 2006). Under-confidence has been theorised to be limiting to motivation (Bandura, 1997), and under-confident students have indeed held lower attitudes towards mathematics than other students, despite having similar or higher performance (for example: Gonida & Leondari, 2011; Narciss et al., 2011; Sheldrake et al., 2014).

Nevertheless, relatively few studies have considered under-confidence and over-confidence, and more research is necessary to gain greater understanding. Earlier studies have also considered varying ages of students and different contexts; more research remains necessary in order to gain greater clarity, and to gain local insights while considering and/or enhancing wider generalisations. Many earlier studies have also considered differences across groups or clusters of students, and further or particular insights may be revealed through considering associations through correlations or predictive models. Clarifying which aspects of education and wider life (including indicators of well-being) associate with under-confidence/over-confidence could lead to valuable educational implications through highlighting areas for awareness and/or support, especially as under-confidence is theorised to be limiting. Students' well-being is also receiving increasing attention within educational research and policies (OECD, 2015a); for example, across Germany, school grades have tended to positively associate with life satisfaction, and further insights will be informative for policy and practice (Autorengruppe Bildungsberichterstattung, 2022).

Accordingly, the presented research aimed to provide new insights into under-confidence and over-confidence by considering a longitudinal sample of secondary school students in Germany. The study considered mathematics in order to increase potential comparability against prior studies and due to its relevance within education and life (OECD, 2015c). The study aimed to reveal what factors predictively associated with under-confidence/over-confidence at different times, and whether/how under-confidence/over-confidence predictively associated with different outcomes.

2. Methods

2.1. Sample

In Germany, the education system can vary across the federal states (Länder). Generally, students study in elementary school from Grade 1 to Grade 4 (age 6/7 to 9/10 years old), secondary school from Grade 5 to Grade 9 (age 10/11 to 14/15 years old), and upper-secondary school from Grade 10 to Grade 12 or Grade 13 (age 15/16 to 18/19 years old), after which students can study at university (if they gain an Abitur certificate at the end of upper-secondary school). Education at secondary school occurs within tracks, usually provided by different types of school: Hauptschulen provide a basic education (usually until Grade 9 or 10); Realschulen provide a more extensive education (usually until Grade 10); Gymnasien provide an intensive education and prepare students for the Abitur. Other schools (such as Gesamtschulen) also offer multiple tracks, while further schools focus on students with special educational needs. The Conference of Ministers of Education has established common educational standards across Länder, focusing on subject-related competencies including mathematics that cover school leaving certificates at Grade 9/10 (KMK, 2022a, 2022b).

This study considered students from the 'Starting Cohort Grade 5' (SC5) in the National Educational Panel Study (NEPS) (Blossfeld & Roßbach, 2019; Blossfeld, Roßbach & von Maurice, 2011; Blossfeld et al., 2011; NEPS, 2019). NEPS sampled schools in Germany (including special educational needs schools) within various strata (including school type) and across multiple Länder, and entire classes of students were invited to participate. Informed consent to participate was gained from the children and their parents (Meixner et al., 2011). Students are surveyed each year, and applying 'sample weighting' allow the students to be representative of the wider population of students across Germany, while accounting for any attrition occurring over time (Aßmann et al., 2011).

The students completed mathematics 'competency' tests in Grades 5 and Grade 9. On each occasion, the students provided indicators of their confidence in their test performance, from which their under-confidence or over-confidence could be calculated. Accordingly, this study focused on 3203 students where this information was available: this excluded students from special educational needs schools (who did not complete any competency tests), excluded students who entered the study part-way through secondary school (via a 'refreshment sample' who did not participate during Grade 5), and also excluded students who had withdrawn from NEPS during this period.

As a simple summary of unweighted numbers and percentages, the 3203 students included 1624 (50.7%) from Gymnasium schools, 735 (22.9%) from Realschulen, 288 (9.0%) from Hauptschulen, with the other 556 students from schools covering various integrated and/or multiple courses of education. At Grade 5, 1584 (49.5%) students identified as girls and 1590 (49.6%) as boys, with the remaining 29 (0.9%) not answering the question (49.9% girls and 50.1% boys when not considering missing information). At Grade 9, 1569 (49.0%) students identified as girls and 1565 (48.9%) as boys, with the remaining 69 (2.2%) not answering the question (50.1% girls and 49.9% boys when not considering missing information). The questionnaires within NEPS have not facilitated students to convey non-binary or other identities.

With sample weighting applied, the sample reflected: 41.5% from Gymnasien, 24.9% from Realschulen, 9.4% from Hauptschulen, with the remaining 24.2% from other schools. At Grade 5, 50.7% identified as girls and 49.3% identified as boys, while at Grade 9, 51.0% identified as girls and 49.0% identified as boys (when not considering missing information).

2.2. Student questionnaires

2.2.1. Items/factors

Questionnaires measured the students' views about various aspects of their education and life, such as their subject-level confidence in mathematics (self-concept beliefs), their well-being considered through self-esteem and life satisfaction, and their home circumstances such as the number of books at home. The number of books at home is often used within educational research (including international studies such as the Programme for International Student Assessment; OECD, 2013, 2015a) as a simple/efficient indicator that may reflect socio-cultural advantage/capital (such as family/guardian resources to purchase books, where books may also reflect 'cultural capital').

Not every indicator was measured each year, however. The analysis therefore considered a specific set of indicators that were measured at Grade 5 and also at Grade 9. This included indicators of the students' subjective well-being, which have received less focus within confidence studies, and so offered the potential for new insights. Numerous other areas were measured by the questionnaires (e.g. reading comprehension strategies, subject-specific interest, etc.), but at intermittent or other times (e.g. at Grade 5 only, from Grade 7 onwards, etc.).

Some indicators were single-item while others were multiple-item factors. Factors were calculated as the average of the relevant questionnaire items, reverse-scoring items when necessary to maintain consistency (so that higher response values/scores then consistently reflected higher attitudes, beliefs, grades, etc.). This process ensured that the factor values remained on the same underlying scale as the questionnaire items, for example where the mathematics self-concept scale was on the same (1) 'does not apply at all' to (4) 'applies completely' scale as the underlying questionnaire items. The theorised assignment of items to factors was verified through factor analysis, and the factors showed acceptable reliability (internal consistency) measured via Cronbach's alpha coefficients.

The analysis considered the following indicators.

- Students' gender (self-identification as male or female).
- Students' school type (Gymnasien, Realschulen, Hauptschulen or other school types), from the initial sample stratum (operationalised as unchanging over time).
- The number of books at home (a measure of socio-economic circumstances), measured as (1) 'none or very few (0 to 10 books)', (2) 'enough to fill one shelf (11 to 25 books)', (3) 'enough to fill several shelves (26 to 100 books)', (4) 'enough to fill a small set of shelves (101 to 200 books)', (5) 'enough to fill a large set of shelves (201 to 500 books)', and (6) 'enough to fill a shelf unit (more than 500 books)'.
- Mathematics grades (from the students' last annual report) at Grade 5 and at Grade 9, measured as (1) 'failing', (2) 'poor', (3) 'passing', (4) 'satisfactory', (5) 'good', and (6) 'very good' grades.
- Mathematics self-concept beliefs (3 items, i.e. 'I get good grades in maths', 'Maths is one of my best subjects', 'I have always been good at maths'; Grade 5 Cronbach's alpha = 0.867, Grade 9 Cronbach's alpha = 0.888), measured from (1) 'does not apply at all', (2) 'does not really apply', (3) 'applies to some extent', and (4) 'applies completely'.
- School self-concept beliefs (3 items, i.e. 'I learn fast in most of the school subjects', 'In most of the school subjects, I perform well in written class tests', 'I perform well in most of the school subjects'; Grade 5 Cronbach's alpha = 0.813, Grade 9 Cronbach's alpha = 0.841), measured from (1) 'does not apply at all', (2) 'does not really apply', (3) 'applies to some extent', and (4) 'applies completely'.
- Self-esteem (10 items, e.g. 'All in all, I am satisfied with myself', 'I can do many things just as well as most other people'; Grade 5 Cronbach's alpha = 0.793, Grade 9 Cronbach's alpha = 0.897), measured from (1) 'does not apply at all', (2) 'does not really apply', (3) 'partially applies', (4) 'applies to some extent', and (5) 'applies completely'.
- Life satisfaction ('How satisfied are you currently and in general terms?') across various areas, including 'with your life', 'with your health', and 'with school' (6 items; Grade 5 Cronbach's alpha = 0.854, Grade 9 Cronbach's alpha = 0.857), measured from (0) 'completely dissatisfied' to (10) 'completely satisfied'.

2.2.2. Under-confidence and over-confidence

Students completed questionnaires each year, but only undertook 'competency' tests in particular years. At Grade 5 and Grade 9, students' mathematics performance was measured via over 20 applied tasks, adapted to students' ages, across content areas of 'quantity', 'change and relationships', 'space and shape', and 'data and chance' (Neumann et al., 2013; Weinert et al., 2011). In the analysis, the mathematics test score reflects the students' actual proportion of correct answers (as a proportion on a 0 to 1 scale).

The students were asked how many mathematics tasks they thought they had answered correctly. In the analysis, mathematics test confidence reflects the students' expected proportion of correct answers (as a proportion on a 0 to 1 scale).

An indicator of under-confidence/over-confidence bias (also referred to as 'calibration') was calculated by comparing the students' proportion of correct answers against the students' estimated proportion of correct answers. Expressed as an equation: mathematics test calibration = mathematics test confidence [the expected proportion correct] – mathematics test score [the actual proportion correct]. Calibration values therefore ranged from -1 (completely under-confident) through 0 (perfectly accurately-evaluated confidence) to +1 (completely over-confident). Essentially, different sides of the scale have different conceptual meaning; this is similar to other scales such as life satisfaction, which range from dissatisfaction to satisfaction. The calibration indicator could not be calculated when students did not have a score and/or expected proportion correct (170 students at Grade 5 and 983 at Grade 9).

Variants of this approach (calibration = expected proportion correct – actual proportion correct) have been frequently applied within research (for example: Bol et al., 2005, 2012; Lichtenstein & Fischhoff, 1977). Conceptually, calibration has also been referred to as 'procedural metacognition' (Lockl, 2013): someone's confidence can be conceptualised as a self-reflective or 'metacognitive' evaluation or judgement, so that calibration then reflects or provides an indicator of theorised processes of self-awareness and 'metacognitive monitoring' during teaching/learning activities such as undertaking mathematics tasks (Pieschl, 2009; Schraw, 2009).

2.3. Analytical approaches

All analysis applied sample-weighting, so that the sample would be representative of the wider population of students across Germany, while accounting for attrition occurring over time (Aßmann et al., 2011).

The profile of the cohort of students was considered through descriptive statistics, and observed associations between the various indicators were considered through correlation coefficients. Correlation coefficients below 0.10 are often considered as minimal associations, from 0.10 to 0.30 as small, from 0.30 to 0.50 as medium/moderate, and above 0.50 as large/strong (Cohen, 1988). Observed differences across gender and across Grade 5 and Grade 9 were quantified through Cohen's D values, where magnitudes below 0.20 are often considered as minimal, from 0.20 to 0.50 as small, from 0.50 to 0.80 as medium/moderate, and above 0.80 as large (Cohen, 1988).

Preliminary analysis used sample-weighted ordinary-least-squares linear regression and sample-weighted linear multi-level modelling, and affirmed that models showed acceptable fit to the data and that various assumptions of linear modelling were met. For example, there were no indicators of collinearity; the average Variance Inflation Factor value per predictor was 1.26 (ranging from 1.00 to 2.02, varying per predictor across the various models; a value above 5 may suggest collinearity where a predictor may offer little additional information to a model), excepting when modelling Grade 5 gender and Grade 9 gender as concurrent predictors. Further analysis then affirmed that findings were the same regardless of whether the relevant model considered gender reported at Grade 5, Grade 9, or both (see the appended Supplementary Material). Overall, the same core findings emerged across ordinary-least-squares regression and multi-level modelling.

The final predictive modelling was then undertaken through sample-weighted linear multi-level modelling (also called mixed modelling) using maximum-likelihood estimation, which accounted for students being clustered with schools (Snijders & Bosker, 2012). This modelling essentially accounted for students potentially having some similarities to each other within schools, given that students experience the same school environment and teaching. In contrast, ordinary-least-squares regression assumes that students are completely independent to each other. Multi-level modelling provides an indication of the amount of variance in the outcome that is explained by the predictors within a model, which is analogous to an adjusted R^2 value produced through ordinary-least-squares regression.

The predictive modelling revealed the independent association between each predictor (such as mathematics self-concept beliefs) and the outcome (such as the calibration indicator), while accounting for all of the other predictors. The standardised predictive coefficients show the number of standard deviations of increase/decrease in the outcome, given one standard deviation increase in the predictor (and can therefore be compared across the different predictors even when the predictors have different scales and units). There are no established thresholds or standards for interpreting magnitudes of standardised predictive coefficients.

Predictive modelling was undertaken in steps. Initially, indicators of students' background and circumstances were included as predictors; then mathematics test calibration bias was also included as a predictor (if calibration bias was not the outcome being predicted); then the remaining indicators were also included as predictors. This process also clarified how much variance was explained by different arrays of predictors.

2.3.1. Cross-sectional predictive modelling

Predictive modelling was applied to consider cross-sectional perspectives at Grade 5 and at Grade 9. This considered the predictors and outcomes at the same time point, for example calibration bias at Grade 5 was predicted by other aspects of education and life at Grade 5. The models considered the following as dependant (outcome) variables:

- Mathematics test calibration bias;
- Mathematics grades;
- Mathematics self-concept beliefs;
- School self-concept beliefs;
- Self-esteem;
- Life satisfaction.

These indicators were also used as independent variables (predictors) when they were not used as dependent variables (outcomes). For example, one model predicted Grade 5 mathematics test calibration bias using Grade 5 mathematics grades, Grade 5 mathematics self-concept beliefs, and an array of other indictors. Another model then predicted Grade 5 mathematics grades using Grade 5 mathematics test calibration bias, Grade 5 mathematics self-concept beliefs, and an array of other indictors. This allowed the modelling to reveal any potentially reciprocal associations. Existing research has highlighted that confidence and performance (for example) can closely associate: for students at secondary school in Germany, mathematics grades have predicted mathematics confidence while mathematics confidence has also predicted mathematics grades (Möller et al., 2011).

Predictive modelling did not include the test score and test confidence indicators, because the calibration bias indicator was calculated from these; the calibration bias indicator can, in theory, be perfectly predicted by the test score and test confidence indicators, such that predictive models cannot then converge on solutions. The analysis instead focused on clarifying how the calibration bias indicator was predicted by (and/or could predict) a range of outcomes covering many aspects of education and life.

2.3.2. Longitudinal predictive modelling

For further insight, given the main focus on calibration bias, a longitudinal model considered predictors of calibration bias at Grade 9, including calibration bias at Grade 5 and all of the other indicators measured at Grade 5 and at Grade 9 as potential predictors. This longitudinal model essentially considered which aspects of education and life could predict the students' calibration bias at Grade 9, over and above the students' earlier calibration bias as of Grade 5, in order to consider how/why changes over time might occur.

3. Results

3.1. Sample profile

On average across the sample, the students were somewhat accurate but tended towards over-confidence (Table 1a): the average mathematics task calibration was 0.23 at Grade 5 and 0.20 at Grade 9. On average, the observed differences and correlations across Grade 5 and Grade 9 tended to be small to moderate (Table 1b). There was no difference in mathematics task calibration across boys and girls at Grade 5 (Table 1c) but there was a difference at Grade 9 (Table 1d); at Grade 9, on average, boys reported greater over-confidence than girls.

3.2. Observed correlations

The mathematics test scores (the proportion of correct answers), test confidence (the proportion of mathematics tasks the student thought they had answered correctly), and calibration bias (the indicator of under-confidence/over-confidence) will correlate because the calibration bias indicator was calculated from the score and confidence indicators.

At Grade 5 (Table 2a), the calibration bias indicator had small positive correlations with students studying in Hauptschulen and Realschulen. The calibration bias indicator also had medium negative correlations with mathematics grades and students studying in Gymnasien, and small negative correlations with numbers of books at home and self-esteem. The calibration bias indicator was scaled from -1 reflecting under-confidence through +1 reflecting over-confidence. The correlations show, for example, higher mathematics grades associating with lower calibration bias (where lower calibration bias means calibration bias towards the direction of under-confidence). This may not necessarily mean that those with higher grades are under-confident, and could mean that those with higher grades have lower extents of over-confidence than others.

At Grade 9 (Table 2b), the calibration bias indicator had small positive correlations with studying in Hauptschulen, mathematics self-concept beliefs, being a boy, and self-esteem. The calibration bias indicator also had small negative correlations with studying in Gymnasien and numbers of books at home.

The wider patterns of correlations revealed various other insights. For example, subject grades, and subject self-concept beliefs and school self-concept beliefs, all positively correlated at Grade 5 and at Grade 9. At Grade 5 and Grade 9, mathematics grades positively correlated with mathematics task scores, while mathematics self-concept positively correlated with mathematics task confidence. Some patterns of correlations were consistently observed across time. For example, at Grade 5 and Grade 9, studying in studying Gymnasien and numbers of books at home negatively correlated with calibration bias. Other correlations differed across time. For example, at Grade 5 self-esteem and life satisfaction negatively correlated with calibration bias, while at Grade 9 self-esteem and life satisfaction positively correlated with calibration bias.

Table 1a
Sample profile (cross-sections).

	Grade 5			Grade 9		
Indicator (scale)	M	SD	N	M	SD	N
Mathematics test score (0-1)	.55	.20	3033	.44	.16	2219
Mathematics test confidence (0-1)	.78	.17	3033	.64	.21	2219
Mathematics test calibration bias (-1 to +1)	.23	.22	3033	.20	.21	2219
Gymnasium (0=no, 1=yes)	.41	.49	3203	.41	.49	3203
Realschule (0=no, 1=yes)	.25	.43	3203	.25	.43	3203
Hauptschule (0=no, 1=yes)	.09	.29	3203	.09	.29	3203
Other school (0=no, 1=yes)	.24	.43	3203	.24	.43	3203
Gender (0=girl, 1=boy)	.49	.50	3169	.49	.50	3095
Books at home (1-6)	3.82	1.41	3016	4.02	1.53	3095
Mathematics grades (1-6)	4.75	.89	2980	4.14	1.02	3007
Mathematics self-concept (1-4)	2.93	.85	3116	2.54	.90	3105
School self-concept (1-4)	3.15	.59	3110	2.90	.59	3096
General self-esteem (1-5)	3.95	.66	3167	3.82	.71	3111
General life satisfaction (0-10)	8.54	1.69	3122	7.88	1.68	3106

Notes: The sample profile shows a cross-sectional perspective, covering all students with data at Grade 5 (even if data was not available at Grade 9) and all students with data at Grade 9 (even if data was not available at Grade 5). Sample-weighted means ('M'), standard deviations ('SD'), and numbers of students ('N') are reported.

Table 1b
Sample profile (paired differences).

	Grade 5			Grade 9			Magnitude of difference between Grade 5 and Grade 9		Correlation between Grade 5 and Grade 9	
Indicator (scale)	M	SD	N	M	SD	N	D	Sig. (p)	R	Sig. (p)
Mathematics test score (0-1)	.55	.20	2084	.44	.15	2084	**.663**	**<.001**	**.576**	**<.001**
Mathematics test confidence (0-1)	.78	.17	2084	.64	.21	2084	**.585**	**<.001**	**.224**	**<.001**
Mathematics test calibration bias (-1 to +1)	.23	.21	2084	.20	.21	2084	**.101**	**<.001**	**.189**	**<.001**
Gender (0=girl, 1=boy)	.49	.50	3062	.49	.50	3062	.018	.308	**.988**	**<.001**
Books at home (1-6)	3.83	1.40	2918	4.05	1.51	2918	**.164**	**<.001**	**.594**	**<.001**
Mathematics grades (1-6)	4.76	.88	2808	4.14	1.02	2808	**.540**	**<.001**	**.292**	**<.001**
Mathematics self-concept (1-4)	2.93	.85	3029	2.53	.90	3029	**.425**	**<.001**	**.430**	**<.001**
School self-concept (1-4)	3.15	.59	3016	2.90	.59	3016	**.354**	**<.001**	**.272**	**<.001**
General self-esteem (1-5)	3.95	.66	3076	3.82	.71	3076	**.161**	**<.001**	**.319**	**<.001**
General life satisfaction (0-10)	8.54	1.70	3027	7.88	1.68	3027	**.312**	**<.001**	**.224**	**<.001**

Notes: The sample profile shows a longitudinal perspective, only covering students with data at Grade 5 and data at Grade 9 (omitting students with data at only one time point). Sample-weighted means ('M'), standard deviations ('SD'), numbers of students ('N'), magnitudes of difference ('D' showing Cohen's D), correlations coefficients ('R'), and significance ('Sig. (p)') are reported. Significant magnitudes/coefficients and significance values (p < .05) are highlighted in bold. The students' school type (Gymnasien, Realschulen, Hauptschulen or other school types) followed from the initial sample stratum and was operationalised as unchanging over time.

Table 1c
Sample profile (gender differences at Grade 5).

	Grade 5 girls			Grade 5 boys			Magnitude of difference between genders	
Indicator (scale)	M	SD	N	M	SD	N	D	Sig. (p)
Grade 5 Mathematics test score (0-1)	.52	.20	1510	.58	.20	1489	**.261**	**<.001**
Grade 5 Mathematics test confidence (0-1)	.76	.17	1510	.81	.16	1489	**.282**	**<.001**
Grade 5 Mathematics test calibration bias (-1 to +1)	.24	.23	1510	.23	.21	1489	.017	.632
Gymnasium (0=no, 1=yes)	.43	.50	1605	.40	.49	1564	.055	.123
Realschule (0=no, 1=yes)	.25	.43	1605	.25	.43	1564	.009	.800
Hauptschule (0=no, 1=yes)	.08	.27	1605	.10	.31	1564	**.080**	**.024**
Other school (0=no, 1=yes)	.24	.43	1605	.24	.43	1564	.000	.993
Grade 5 Books at home (1-6)	3.86	1.34	1528	3.78	1.48	1470	.062	.087
Grade 5 Mathematics grades (1-6)	4.62	.90	1520	4.88	.86	1448	**.306**	**<.001**
Grade 5 Mathematics self-concept (1-4)	2.70	.85	1587	3.18	.78	1511	**.590**	**<.001**
Grade 5 School self-concept (1-4)	3.14	.59	1585	3.15	.59	1507	.019	.596
Grade 5 General self-esteem (1-5)	3.94	.67	1601	3.96	.65	1549	.028	.427
Grade 5 General life satisfaction (0-10)	8.68	1.55	1578	8.39	1.81	1528	**.172**	**<.001**

Notes: Sample-weighted means ('M'), standard deviations ('SD'), numbers of students ('N'), and magnitudes of difference ('D' showing Cohen's D) and significance ('Sig. (p)') are reported. Significant magnitudes and significance values ($p < .05$) are highlighted in bold.

Table 1d
Sample profile (gender differences at Grade 9).

	Grade 9 girls			Grade 9 boys			Magnitude of difference between genders	
Indicator (scale)	M	SD	N	M	SD	N	D	Sig. (p)
Grade 9 Mathematics test score (0-1)	.42	.15	1129	.45	.16	1072	**.207**	**<.001**
Grade 9 Mathematics test confidence (0-1)	.60	.20	1129	.69	.20	1072	**.441**	**<.001**
Grade 9 Mathematics test calibration bias (-1 to +1)	.18	.21	1129	.23	.21	1072	**.272**	**<.001**
Gymnasium (0=no, 1=yes)	.42	.49	1577	.40	.49	1518	.040	.259
Realschule (0=no, 1=yes)	.26	.44	1577	.25	.43	1518	.017	.643
Hauptschule (0=no, 1=yes)	.08	.27	1577	.11	.31	1518	**.089**	**.013**
Other school (0=no, 1=yes)	.24	.43	1577	.24	.43	1518	.003	.927
Grade 9 Books at home (1-6)	4.12	1.46	1573	3.92	1.58	1502	**.134**	**<.001**
Grade 9 Mathematics grades (1-6)	4.14	1.00	1541	4.15	1.03	1447	.013	.728
Grade 9 Mathematics self-concept (1-4)	2.35	.89	1576	2.74	.87	1509	**.441**	**<.001**
Grade 9 School self-concept (1-4)	2.89	.61	1569	2.91	.56	1507	.029	.416
Grade 9 General self-esteem (1-5)	3.66	.71	1577	3.99	.68	1514	**.474**	**<.001**
Grade 9 General life satisfaction (0-10)	7.77	1.73	1575	8.00	1.61	1513	**.134**	**<.001**

Notes: Sample-weighted means ('M'), standard deviations ('SD'), numbers of students ('N'), and magnitudes of difference ('D' showing Cohen's D) and significance ('Sig. (p)') are reported. Significant magnitudes and significance values ($p < .05$) are highlighted in bold.

Table 2a
Correlations at Grade 5.

Indicator (scale)	1a	1b	1c	2a	2b	2c	2d	3	4	5	6	7	8
[1a] Mathematics test score (0-1)	**1**												
[1b] Mathematics test confidence (0-1)	**.312**	**1**											
[1c] Mathematics test calibration bias (-1 to +1)	**-.664**	**.503**	**1**										
[2a] Gymnasium (0=no, 1=yes)	**.455**	**.131**	**-.311**	**1**									
[2b] Realschule (0=no, 1=yes)	**-.144**	**-.044**	**.097**	**-.485**	**1**								
[2c] Hauptschule (0=no, 1=yes)	**-.311**	**-.078**	**.221**	**-.271**	**-.185**	**1**							
[2d] Other school (0=no, 1=yes)	**-.173**	**-.054**	**.114**	**-.476**	**-.326**	**-.182**	**1**						
[3] Gender (0=girl, 1=boy)	**.129**	**.140**	-.009	-.027	.005	**.040**	.000	**1**					
[4] Books at home (1-6)	**.344**	**.097**	**-.236**	**.303**	**-.130**	**-.148**	**-.123**	-.031	**1**				
[5] Mathematics grades (1-6)	**.540**	**.231**	**-.313**	**.386**	**-.188**	**-.297**	**-.063**	**.151**	**.265**	**1**			
[6] Mathematics self-concept (1-4)	**.322**	**.264**	**-.088**	**.092**	-.030	**-.062**	-.035	**.283**	**.105**	**.479**	**1**		
[7] School self-concept (1-4)	**.187**	**.120**	**-.077**	**.117**	**-.083**	-.014	**-.042**	.010	**.189**	**.254**	**.340**	**1**	
[8] General self-esteem (1-5)	**.186**	**.077**	**-.110**	**.140**	-.030	**-.101**	**-.062**	.014	**.149**	**.183**	**.185**	**.361**	**1**
[9] General life satisfaction (0-10)	**.096**	-.002	**-.090**	**.137**	**-.037**	**-.099**	**-.054**	**-.086**	**.096**	**.119**	**.112**	**.282**	**.411**

Notes: Sample-weighted correlations coefficients (R) are reported. Significant coefficients ($p < .05$) are highlighted in bold.

Table 2b
Correlations at Grade 9.

Indicator (scale)	1a	1b	1c	2a	2b	2c	2d	3	4	5	6	7	8
[1a] Mathematics test score (0-1)	**1**												
[1b] Mathematics test confidence (0-1)	**.354**	**1**											
[1c] Mathematics test calibration bias (-1 to +1)	**-.396**	**.719**	**1**										
[2a] Gymnasium (0=no, 1=yes)	**.423**	**.158**	**-.160**	**1**									
[2b] Realschule (0=no, 1=yes)	**-.125**	**-.111**	-.017	**-.485**	**1**								
[2c] Hauptschule (0=no, 1=yes)	**-.239**	.002	**.180**	**-.271**	**-.185**	**1**							
[2d] Other school (0=no, 1=yes)	**-.220**	**-.074**	**.091**	**-.476**	**-.326**	**-.182**	**1**						
[3] Gender (0=girl, 1=boy)	**.103**	**.215**	**.135**	-.020	-.008	**.044**	.002	**1**					
[4] Books at home (1-6)	**.344**	**.098**	**-.160**	**.302**	**-.134**	**-.208**	**-.072**	**-.067**	**1**				
[5] Mathematics grades (1-6)	**.323**	**.265**	.021	**.062**	**-.062**	-.025	.007	.006	**.209**	**1**			
[6] Mathematics self-concept (1-4)	**.244**	**.365**	**.177**	**.040**	-.020	.005	-.029	**.215**	**.090**	**.603**	**1**		
[7] School self-concept (1-4)	**.194**	**.241**	**.092**	**.089**	**-.064**	-.026	-.020	.015	**.182**	**.368**	**.353**	**1**	
[8] General self-esteem (1-5)	**.141**	**.213**	**.104**	**.053**	**-.067**	-.007	.012	**.231**	**.135**	**.142**	**.218**	**.356**	**1**
[9] General life satisfaction (0-10)	**.082**	**.139**	**.075**	**.050**	**-.055**	.010	-.009	**.067**	**.110**	**.173**	**.164**	**.344**	**.499**

Notes: Sample-weighted correlations coefficients (R) are reported. Significant coefficients ($p < .05$) are highlighted in bold.

3.3. *Cross-sectional predictive modelling*

3.3.1. *Predicting mathematics test calibration bias*

Calibration bias at Grade 5 (Table 3a) was positively predicted by studying in 'other' school types, in Hauptschulen, and in

Table 3a
Predicting mathematics test calibration bias at Grade 5 using Grade 5 predictors.

Predictor	Step 1 Std.	Sig. (p)	Step 2 Std.	Sig. (p)
Intercept	**N/A**	**<.001**	**N/A**	**<.001**
Realschule (1=yes, compared to Gymnasium)	**.234**	**<.001**	**.180**	**<.001**
Hauptschule (1=yes, compared to Gymnasium)	**.286**	**<.001**	**.217**	**<.001**
Other school (1=yes, compared to Gymnasium)	**.262**	**<.001**	**.227**	**<.001**
Grade 5 Gender (1=boy, compared to 0=girl)	**-.045**	**.009**	-.024	.187
Grade 5 Books at home	**-.128**	**<.001**	**-.106**	**<.001**
Grade 5 Mathematics grades			**-.172**	**<.001**
Grade 5 Mathematics self-concept			.024	.263
Grade 5 School self-concept			-.014	.477
Grade 5 General self-esteem			-.006	.778
Grade 5 General life satisfaction			-.018	.354
Explained variance	18.6%		21.2%	
Unexplained variance (residual-level)	70.0%		68.5%	
Unexplained variance (school-level)	11.4%		10.3%	

Notes: Sample-weighted standardised predictive coefficients ('Std.') and significance ('Sig. (p)') are reported. Significant coefficients and significance values ($p < .05$) are highlighted in bold.

Realschulen (all compared to studying in Gymnasien). Calibration bias at Grade 5 was also negatively predicted by mathematics grades at Grade 5 and by numbers of books at home at Grade 5.

Calibration bias at Grade 9 (Table 3b) was positively predicted by mathematics self-concept at Grade 9, by studying in Hauptschulen and in 'other' school types (both compared to studying in Gymnasien), by self-esteem, by school self-concept, and by being a boy (compared to being a girl). Calibration bias at Grade 9 was also negatively predicted by mathematics grades and by numbers of books at home.

3.3.2. Predicting other outcomes

Further modelling predicted other outcomes at Grade 5 and Grade 9: mathematics grades (Tables 4a and 4b), mathematics self-concept (Tables 5a and 5b), school self-concept (Tables 6a and 6b), self-esteem (Tables 7a and 7b), and life satisfaction (Tables 8a and 8b).

At Grade 5, while accounting for all of the other considered predictors, calibration bias negatively predicted mathematics grades (Table 4a). Nevertheless, at Grade 5, calibration bias did not significantly predict mathematics self-concept, school self-concept, self-esteem, and life satisfaction.

At Grade 9, while accounting for all of the other considered predictors, calibration bias negatively predicted mathematics grades (Table 4b). At Grade 9, calibration bias also positively predicted mathematics self-concept (Table 5b), school self-concept (Table 6b), and self-esteem (Table 7b), while accounting for the other predictors. Nevertheless, calibration bias did not significantly predict life satisfaction.

3.4. Longitudinal predictive modelling

Calibration bias at Grade 9 (Table 9) was positively predicted by mathematics self-concept at Grade 9, by calibration bias at Grade 5, by studying in Hauptschulen (compared to studying in Gymnasien), by school self-concept at Grade 9, by studying in 'other' school types (compared to studying in Gymnasien), by mathematics self-concept at Grade 5, and by self-esteem at Grade 9. Calibration bias at Grade 9 was also negatively predicted by school self-concept at Grade 5 and by mathematics grades at Grade 9.

4. Discussion

These findings offer various new insights to help understand under-confidence and over-confidence within the context of students' education and lives. The analysis considered a representative cohort of students in Germany over time, from Grade 5 to Grade 9 (age 10/11 to 14/15 years old), and considered what associated with the students' extent of under-confidence or over-confidence (calibration bias) for mathematics tasks through comparing their performance against their own estimates of their performance.

The cross-sectional predictive modelling revealed that, at Grade 5 and also at Grade 9, mathematics calibration bias was negatively predicted by subject-level mathematics grades. Essentially, higher grades predicted lower calibration bias (i.e. higher grades predicted bias towards the direction of under-confidence). At Grade 9, mathematics calibration bias was also positively predicted by mathematics self-concept beliefs, which reflected subject-level confidence measured through questionnaire items including 'I get good grades in maths' and 'Maths is one of my best subjects'. Essentially, higher subject-level mathematics confidence predicted higher calibration bias (towards the direction of over-confidence). The longitudinal predictive modelling revealed that calibration bias at Grade 9 was positively predicted by calibration bias at Grade 5, mathematics self-concept at Grade 9, and mathematics self-concept at Grade 5, alongside various other predictors. These findings help clarify and extend existing understanding through revealing that calibration bias (the extent of under-confidence/over-confidence) on particular mathematics tasks linked with wider subject-level performance

Table 3b
Predicting mathematics test calibration bias at Grade 9 using Grade 9 predictors.

Predictor	Step 1 Std.	Sig. (p)	Step 2 Std.	Sig. (p)
Intercept	**N/A**	**<.001**	N/A	.150
Realschule (1=yes, compared to Gymnasium)	.039	.241	.037	.231
Hauptschule (1=yes, compared to Gymnasium)	**.171**	**<.001**	**.167**	**<.001**
Other school (1=yes, compared to Gymnasium)	**.127**	**.001**	**.134**	**<.001**
Grade 9 Gender (1=boy, compared to 0=girl)	**.099**	**<.001**	**.042**	**.029**
Grade 9 Books at home	**-.058**	**.005**	**-.078**	**<.001**
Grade 9 Mathematics grades			**-.106**	**<.001**
Grade 9 Mathematics self-concept			**.204**	**<.001**
Grade 9 School self-concept			**.047**	**.030**
Grade 9 General self-esteem			**.048**	**.035**
Grade 9 General life satisfaction			.038	.099
Explained variance	7.3%		12.4%	
Unexplained variance (residual-level)	83.9%		80.6%	
Unexplained variance (school-level)	8.8%		7.0%	

Notes: Sample-weighted standardised predictive coefficients ('Std.') and significance ('Sig. (p)') are reported. Significant coefficients and significance values ($p < .05$) are highlighted in bold.

Table 4a
Predicting mathematics grades at Grade 5 using Grade 5 predictors.

Predictor	Step 1 Std.	Sig. (p)	Step 2 Std.	Sig. (p)	Step 3 Std.	Sig. (p)
Intercept	**N/A**	**<.001**	**N/A**	**<.001**	**N/A**	**<.001**
Grade 5 Mathematics test calibration bias			**-.144**	**<.001**	**-.118**	**<.001**
Realschule (1=yes, compared to Gymnasium)	**-.301**	**<.001**	**-.267**	**<.001**	**-.255**	**<.001**
Hauptschule (1=yes, compared to Gymnasium)	**-.397**	**<.001**	**-.355**	**<.001**	**-.332**	**<.001**
Other school (1=yes, compared to Gymnasium)	**-.187**	**<.001**	**-.150**	**<.001**	**-.139**	**<.001**
Grade 5 Gender (1=boy, compared to 0=girl)	**.170**	**<.001**	**.163**	**<.001**	**.061**	**<.001**
Grade 5 Books at home	**.125**	**<.001**	**.106**	**<.001**	**.062**	**<.001**
Grade 5 Mathematics self-concept					**.377**	**<.001**
Grade 5 School self-concept					**.046**	**.006**
Grade 5 General self-esteem					.016	.341
Grade 5 General life satisfaction					.006	.709
Explained variance	28.5%		30.5%		44.0%	
Unexplained variance (residual-level)	61.7%		60.3%		47.9%	
Unexplained variance (school-level)	9.8%		9.1%		8.1%	

Notes: Sample-weighted standardised predictive coefficients ('Std.') and significance ('Sig. (p)') are reported. Significant coefficients and significance values ($p < .05$) are highlighted in bold.

Table 4b
Predicting mathematics grades at Grade 9 using Grade 9 predictors.

Predictor	Step 1 Std.	Sig. (p)	Step 2 Std.	Sig. (p)	Step 3 Std.	Sig. (p)
Intercept	**N/A**	**<.001**	**N/A**	**<.001**	**N/A**	**<.001**
Grade 9 Mathematics test calibration bias			**.058**	**.004**	**-.065**	**<.001**
Realschule (1=yes, compared to Gymnasium)	**-.061**	**.047**	**-.063**	**.042**	**-.051**	**.047**
Hauptschule (1=yes, compared to Gymnasium)	.004	.893	-.006	.821	-.005	.813
Other school (1=yes, compared to Gymnasium)	-.001	.968	-.009	.801	.019	.503
Grade 9 Gender (1=boy, compared to 0=girl)	**.061**	**.002**	**.055**	**.004**	**-.063**	**<.001**
Grade 9 Books at home	**.188**	**<.001**	**.191**	**<.001**	**.084**	**<.001**
Grade 9 Mathematics self-concept					**.571**	**<.001**
Grade 9 School self-concept					**.173**	**<.001**
Grade 9 General self-esteem					**-.053**	**.003**
Grade 9 General life satisfaction					**.076**	**<.001**
Explained variance	5.6%		5.6%		45.6%	
Unexplained variance (residual-level)	88.2%		87.8%		49.1%	
Unexplained variance (school-level)	6.2%		6.6%		5.3%	

Notes: Sample-weighted standardised predictive coefficients ('Std.') and significance ('Sig. (p)') are reported. Significant coefficients and significance values ($p < .05$) are highlighted in bold.

Table 5a
Predicting mathematics self-concept at Grade 5 using Grade 5 predictors.

Predictor	Step 1 Std.	Sig. (p)	Step 2 Std.	Sig. (p)	Step 3 Std.	Sig. (p)
Intercept	**N/A**	**<.001**	**N/A**	**<.001**	**N/A**	**<.001**
Grade 5 Mathematics test calibration bias			**-.055**	**.006**	.021	.233
Realschule (1=yes, compared to Gymnasium)	-.038	.221	-.026	.420	**.103**	**.001**
Hauptschule (1=yes, compared to Gymnasium)	**-.077**	**.005**	**-.062**	**.029**	**.086**	**.001**
Other school (1=yes, compared to Gymnasium)	-.038	.210	-.024	.442	.052	.068
Grade 5 Gender (1=boy, compared to 0=girl)	**.276**	**<.001**	**.274**	**<.001**	**.202**	**<.001**
Grade 5 Books at home	**.098**	**<.001**	**.091**	**<.001**	<.001	.984
Grade 5 Mathematics grades					**.429**	**<.001**
Grade 5 School self-concept					**.248**	**<.001**
Grade 5 General self-esteem					.035	.051
Grade 5 General life satisfaction					.005	.762
Explained variance	10.1%		10.2%		33.6%	
Unexplained variance (residual-level)	82.0%		81.7%		59.0%	
Unexplained variance (school-level)	7.9%		8.1%		7.4%	

Notes: Sample-weighted standardised predictive coefficients ('Std.') and significance ('Sig. (p)') are reported. Significant coefficients and significance values ($p < .05$) are highlighted in bold.

Table 5b
Predicting mathematics self-concept at Grade 9 using Grade 9 predictors.

Predictor	Step 1 Std.	Sig. (p)	Step 2 Std.	Sig. (p)	Step 3 Std.	Sig. (p)
Intercept	**N/A**	**<.001**	**N/A**	**<.001**	**N/A**	**<.001**
Grade 9 Mathematics test calibration bias			**.183**	**<.001**	**.134**	**<.001**
Realschule (1=yes, compared to Gymnasium)	-.013	.613	-.020	.443	.021	.350
Hauptschule (1=yes, compared to Gymnasium)	.017	.508	-.014	.567	-.010	.621
Other school (1=yes, compared to Gymnasium)	-.009	.749	-.031	.263	-.021	.394
Grade 9 Gender (1=boy, compared to 0=girl)	**.226**	**<.001**	**.208**	**<.001**	**.164**	**<.001**
Grade 9 Books at home	**.105**	**<.001**	**.117**	**<.001**	-.022	.181
Grade 9 Mathematics grades					**.589**	**<.001**
Grade 9 School self-concept					**.110**	**<.001**
Grade 9 General self-esteem					**.047**	**.010**
Grade 9 General life satisfaction					-.025	.170
Explained variance	5.7%		8.7%		45.8%	
Unexplained variance (residual-level)	91.3%		88.7%		51.3%	
Unexplained variance (school-level)	3.0%		2.6%		2.9%	

Notes: Sample-weighted standardised predictive coefficients ('Std.') and significance ('Sig. (p)') are reported. Significant coefficients and significance values ($p < .05$) are highlighted in bold.

Table 6a
Predicting school self-concept at Grade 5 using Grade 5 predictors.

Predictor	Step 1 Std.	Sig. (p)	Step 2 Std.	Sig. (p)	Step 3 Std.	Sig. (p)
Intercept	**N/A**	**<.001**	**N/A**	**<.001**	**N/A**	**<.001**
Grade 5 Mathematics test calibration bias			**-.050**	**.017**	-.009	.607
Realschule (1=yes, compared to Gymnasium)	-.054	.055	-.043	.133	.002	.949
Hauptschule (1=yes, compared to Gymnasium)	.021	.428	.035	.196	**.111**	**<.001**
Other school (1=yes, compared to Gymnasium)	-.038	.170	-.025	.371	.028	.274
Grade 5 Gender (1=boy, compared to 0=girl)	.008	.659	.006	.744	**-.074**	**<.001**
Grade 5 Books at home	**.170**	**<.001**	**.163**	**<.001**	**.097**	**<.001**
Grade 5 Mathematics grades					**.063**	**.004**
Grade 5 Mathematics self-concept					**.283**	**<.001**
Grade 5 General self-esteem					**.238**	**<.001**
Grade 5 General life satisfaction					**.155**	**<.001**
Explained variance	3.4%		3.6%		26.4%	
Unexplained variance (residual-level)	91.9%		91.6%		69.6%	
Unexplained variance (school-level)	4.7%		4.8%		4.0%	

Notes: Sample-weighted standardised predictive coefficients ('Std.') and significance ('Sig. (p)') are reported. Significant coefficients and significance values ($p < .05$) are highlighted in bold.

Table 6b
Predicting school self-concept at Grade 9 using Grade 9 predictors.

Predictor	Step 1 Std.	Sig. (p)	Step 2 Std.	Sig. (p)	Step 3 Std.	Sig. (p)
Intercept	**N/A**	**<.001**	**N/A**	**<.001**	**N/A**	**<.001**
Grade 9 Mathematics test calibration bias			**.118**	**<.001**	**.043**	**.019**
Realschule (1=yes, compared to Gymnasium)	-.019	.472	-.024	.370	.003	.908
Hauptschule (1=yes, compared to Gymnasium)	.035	.170	.015	.551	.030	.171
Other school (1=yes, compared to Gymnasium)	-.027	.357	-.042	.152	-.015	.541
Grade 9 Gender (1=boy, compared to 0=girl)	**.046**	**.018**	.035	.076	**-.067**	**<.001**
Grade 9 Books at home	**.195**	**<.001**	**.202**	**<.001**	**.100**	**<.001**
Grade 9 Mathematics grades					**.234**	**<.001**
Grade 9 Mathematics self-concept					**.148**	**<.001**
Grade 9 General self-esteem					**.227**	**<.001**
Grade 9 General life satisfaction					**.119**	**<.001**
Explained variance	4.1%		5.4%		28.8%	
Unexplained variance (residual-level)	92.6%		91.6%		69.4%	
Unexplained variance (school-level)	3.3%		3.0%		1.8%	

Notes: Sample-weighted standardised predictive coefficients ('Std.') and significance ('Sig. (p)') are reported. Significant coefficients and significance values ($p < .05$) are highlighted in bold.

Table 7a
Predicting general self-esteem at Grade 5 using Grade 5 predictors.

Predictor	Step 1 Std.	Sig. (p)	Step 2 Std.	Sig. (p)	Step 3 Std.	Sig. (p)
Intercept	**N/A**	**<.001**	**N/A**	**<.001**	**N/A**	**<.001**
Grade 5 Mathematics test calibration bias			-.041	.051	-.008	.655
Realschule (1=yes, compared to Gymnasium)	-.052	.072	-.043	.143	-.004	.870
Hauptschule (1=yes, compared to Gymnasium)	**-.102**	**<.001**	**-.090**	**.001**	**-.057**	**.027**
Other school (1=yes, compared to Gymnasium)	**-.100**	**.001**	**-.090**	**.002**	-.051	.056
Grade 5 Gender (1=boy, compared to 0=girl)	.013	.499	.011	.559	.020	.253
Grade 5 Books at home	**.118**	**<.001**	**.112**	**<.001**	**.049**	**.007**
Grade 5 Mathematics grades					.023	.306
Grade 5 Mathematics self-concept					.039	.067
Grade 5 School self-concept					**.246**	**<.001**
Grade 5 General life satisfaction					**.323**	**<.001**
Explained variance	3.5%		3.7%		25.0%	
Unexplained variance (residual-level)	91.3%		91.2%		70.3%	
Unexplained variance (school-level)	5.2%		5.1%		4.7%	

Notes: Sample-weighted standardised predictive coefficients ('Std.') and significance ('Sig. (p)') are reported. Significant coefficients and significance values ($p < .05$) are highlighted in bold.

Table 7b
Predicting general self-esteem at Grade 9 using Grade 9 predictors.

Predictor	Step 1 Std.	Sig. (p)	Step 2 Std.	Sig. (p)	Step 3 Std.	Sig. (p)
Intercept	**N/A**	**<.001**	**N/A**	**<.001**	**N/A**	**<.001**
Grade 9 Mathematics test calibration bias			**.112**	**<.001**	**.038**	**.028**
Realschule (1=yes, compared to Gymnasium)	-.035	.152	-.038	.107	-.029	.146
Hauptschule (1=yes, compared to Gymnasium)	-.034	.161	**-.053**	**.028**	**-.055**	**.007**
Other school (1=yes, compared to Gymnasium)	-.046	.081	**-.059**	**.024**	-.024	.269
Grade 9 Gender (1=boy, compared to 0=girl)	**.237**	**<.001**	**.226**	**<.001**	**.184**	**<.001**
Grade 9 Books at home	**.108**	**<.001**	**.117**	**<.001**	**.037**	**.034**
Grade 9 Mathematics grades					**-.061**	**.005**
Grade 9 Mathematics self-concept					**.051**	**.018**
Grade 9 School self-concept					**.204**	**<.001**
Grade 9 General life satisfaction					**.473**	**<.001**
Explained variance	7.2%		8.4%		36.6%	
Unexplained variance (residual-level)	91.3%		90.3%		62.4%	
Unexplained variance (school-level)	1.5%		1.3%		1.0%	

Notes: Sample-weighted standardised predictive coefficients ('Std.') and significance ('Sig. (p)') are reported. Significant coefficients and significance values ($p < .05$) are highlighted in bold.

Table 8a
Predicting general life satisfaction at Grade 5 using Grade 5 predictors.

Predictor	Step 1 Std.	Sig. (p)	Step 2 Std.	Sig. (p)	Step 3 Std.	Sig. (p)
Intercept	**N/A**	**<.001**	**N/A**	**<.001**	**N/A**	**<.001**
Grade 5 Mathematics test calibration bias			**-.044**	**.037**	-.020	.308
Realschule (1=yes, compared to Gymnasium)	**-.076**	**.013**	**-.065**	**.031**	-.041	.149
Hauptschule (1=yes, compared to Gymnasium)	**-.111**	**<.001**	**-.099**	**<.001**	**-.070**	**.008**
Other school (1=yes, compared to Gymnasium)	**-.101**	**.001**	**-.089**	**.003**	-.053	.054
Grade 5 Gender (1=boy, compared to 0=girl)	**-.074**	**<.001**	**-.076**	**<.001**	**-.085**	**<.001**
Grade 5 Books at home	**.055**	**.006**	**.049**	**.015**	-.018	.347
Grade 5 Mathematics grades					.010	.679
Grade 5 Mathematics self-concept					.009	.664
Grade 5 School self-concept					**.167**	**<.001**
Grade 5 General self-esteem					**.338**	**<.001**
Explained variance	3.0%		3.2%		21.1%	
Unexplained variance (residual-level)	91.0%		90.9%		73.9%	
Unexplained variance (school-level)	6.0%		5.9%		5.0%	

Notes: Sample-weighted standardised predictive coefficients ('Std.') and significance ('Sig. (p)') are reported. Significant coefficients and significance values ($p < .05$) are highlighted in bold.

Table 8b
Predicting general life satisfaction at Grade 9 using Grade 9 predictors.

Predictor	Step 1 Std.	Sig. (p)	Step 2 Std.	Sig. (p)	Step 3 Std.	Sig. (p)
Intercept	**N/A**	**<.001**	**N/A**	**<.001**	**N/A**	**<.001**
Grade 9 Mathematics test calibration bias			**.093**	**<.001**	.029	.093
Realschule (1=yes, compared to Gymnasium)	-.011	.663	-.014	.567	.011	.577
Hauptschule (1=yes, compared to Gymnasium)	.015	.529	-.001	.975	.022	.278
Other school (1=yes, compared to Gymnasium)	-.043	.103	**-.054**	**.040**	-.022	.308
Grade 9 Gender (1=boy, compared to 0=girl)	**.068**	**<.001**	**.059**	**.002**	**-.049**	**.003**
Grade 9 Books at home	**.088**	**<.001**	**.094**	**<.001**	.005	.786
Grade 9 Mathematics grades					**.088**	**<.001**
Grade 9 Mathematics self-concept					-.027	.207
Grade 9 School self-concept					**.105**	**<.001**
Grade 9 General self-esteem					**.464**	**<.001**
Explained variance	1.6%		2.5%		30.9%	
Unexplained variance (residual-level)	96.2%		95.5%		67.9%	
Unexplained variance (school-level)	2.2%		2.0%		1.2%	

Notes: Sample-weighted standardised predictive coefficients ('Std.') and significance ('Sig. (p)') are reported. Significant coefficients and significance values ($p < .05$) are highlighted in bold.

Table 9
Predicting mathematics task calibration at Grade 9 using Grade 5 and Grade 9 predictors.

Predictor	Step 1 Std.	Sig. (p)	Step 2 Std.	Sig. (p)	Step 3 Std.	Sig. (p)
Intercept	**N/A**	**<.001**	**N/A**	**<.001**	N/A	.616
Grade 5 Mathematics task calibration			**.120**	**<.001**	**.141**	**<.001**
Realschule (1=yes, compared to Gymnasium)	.024	.464	<.001	.988	<.001	.995
Hauptschule (1=yes, compared to Gymnasium)	**.172**	**<.001**	**.131**	**<.001**	**.121**	**<.001**
Other school (1=yes, compared to Gymnasium)	**.123**	**.001**	**.084**	**.025**	**.083**	**.021**
Grade 9 Gender (1=boy, compared to 0=girl)	**.092**	**<.001**	**.100**	**<.001**	.031	.163
Grade 5 Books at home	-.010	.709	-.002	.952	-.009	.715
Grade 9 Books at home	-.014	.609	-.007	.778	-.035	.178
Grade 5 Mathematics grades					.019	.517
Grade 9 Mathematics grades					**-.067**	**.016**
Grade 5 Mathematics self-concept					**.068**	**.014**
Grade 9 Mathematics self-concept					**.149**	**<.001**
Grade 5 School self-concept					**-.069**	**.006**
Grade 9 School self-concept					**.097**	**<.001**
Grade 5 General self-esteem					-.011	.631
Grade 9 General self-esteem					**.057**	**.027**
Grade 5 General life satisfaction					-.002	.950
Grade 9 General life satisfaction					.022	.377
Explained variance	5.0%		6.4%		12.2%	
Unexplained variance (residual-level)	86.8%		86.0%		81.8%	
Unexplained variance (school-level)	8.3%		7.7%		6.0%	

Notes: Sample-weighted standardised predictive coefficients ('Std.') and significance ('Sig. (p)') are reported. Significant coefficients and significance values ($p < .05$) are highlighted in bold.

and beliefs. Essentially, judgements or evaluations related to particular tasks may partially follow from students' general experiences and views, which may help (partially) explain why calibration biases may arise.

The other aspects of the cross-sectional predictive modelling revealed important insights into the potential implications of calibration bias. Specifically, at Grade 5 and at Grade 9, calibration bias negatively predicted mathematics grades. At Grade 9, calibration bias positively predicted mathematics self-concept, school self-concept, and self-esteem, over and above the array of other predictors. The difference between Grade 5 and Grade 9 is an important insight in itself, and suggests that education may need to increasingly consider under-confidence/over-confidence as students grow older; it is possible that the antecedents and/or implications of calibration biases may vary at different ages or stages of education. The longitudinal modelling revealed that calibration bias at Grade 9 was positively predicted by calibration bias at Grade 5, and by mathematics self-concept at Grade 5 and at Grade 9, together with an array of other predictors. These findings offer insight into how and why changes may arise over time, and provide a starting point for more extensive future research. Additionally, considering the averages observed across gender, boys and girls showed similar calibration bias at Grade 5, but boys on average showed higher calibration bias (greater over-confidence) than girls at Grade 9. The predictive modelling then confirmed that being a boy (compared to being a girl) independently predicted higher calibration bias at Grade 9 (accounting for the other predictors). Prior studies have revealed that boys have often exhibited greater over-confidence than girls (Gonida & Leondari, 2011; Sheldrake et al., 2014), and the presented results extend understanding to show that findings may vary at different ages.

Students' well-being is receiving increasing attention within educational research and policies (Autorengruppe Bildungsberichterstattung, 2022; OECD, 2015a). The results presented here revealed new insights: calibration bias had no clear associations with life satisfaction but associated with self-esteem. Specifically, at Grade 9 calibration bias towards the direction of over-confidence predictively associated with higher self-esteem; concurrently, higher self-esteem predictively associated with calibration bias towards the direction of over-confidence. Nevertheless, these results help affirm an association, but cannot show a causal relation (where one is definitively an antecedent of the other); it is also possible that other (unmeasured) aspects of life associate with calibration bias and self-esteem, which may help explain any pattern of observed associations. Relatively few studies with secondary school students have considered students' well-being (such as self-esteem and life satisfaction) in relation to under-confidence/over-confidence. A study of students in Canada (Grades 3/4 to 7/8; age around 9 to 14 years old) showed that students with the highest under-confidence exhibited the lowest self-esteem, while students with the highest over-confidence exhibited the highest self-esteem (Bouffard et al., 2011). Wider psychological studies have produced variable findings: positive associations between over-confidence, self-esteem, and general positive affect have been revealed for adults in Germany (Dufner et al., 2012), for example, while adults in the United States of America with accurate confidence evaluations have reported higher self-esteem than those with over-confident or under-confident evaluations (Kim & Chiu, 2011). The results presented here also revealed that studying in Hauptschulen associated with lower life satisfaction than studying in Gymnasien, and that higher mathematics grades associated with higher life satisfaction; these findings affirm existing research in Germany which has associated grades with life satisfaction (Autorengruppe Bildungsberichterstattung, 2022).

The range of results presented here also offer numerous other insights and implications around calibration and confidence. For example, the students, on average, tended towards over-confidence at Grade 5 and Grade 9. Tendencies towards over-confidence have also been observed within prior studies (Boekaerts & Rozendaal, 2010; Chen, 2003; García et al., 2016). In the results presented here, gender did not associate with calibration bias at Grade 5, but being a boy (compared to a girl) at Grade 9 positively predicted calibration bias at Grade 9. Prior studies have revealed that boys have often exhibited over-confidence while girls have often exhibited under-confidence (for example: Gonida & Leondari, 2011; Sheldrake et al., 2014), while other aspects of students' backgrounds and schools have received less focus. In the results presented here, considering school contexts, studying in Hauptschulen positively correlated with calibration bias while studying in Gymnasien negatively correlated with calibration bias at Grade 5 and at Grade 9; studying in Hauptschulen and 'other' school types (compared to studying in Gymnasien) positively predicted calibration bias at Grade 5 and at Grade 9. Future research may benefit from exploring the implications of schools and other circumstances in more detail.

Finally, the range of results presented here also offer wider insights. Considering the averages observed across boys and girls, at Grade 5, boys and girls showed similar calibration bias, but boys on average showed higher mathematics test scores, mathematics test confidence, mathematics grades, and mathematics self-concept beliefs, but lower life satisfaction. At Grade 9, boys on average showed higher calibration bias and higher mathematics test scores, mathematics test confidence, mathematics self-concept beliefs, self-esteem, and life satisfaction, but lower numbers of books at home. The predictive modelling then revealed, accounting for the other predictors, at Grade 5, being a boy (compared to a girl) predicted higher mathematics grades and higher mathematics self-concept beliefs but lower school self-concept beliefs; at Grade 9, being a boy predicted higher mathematics self-concept beliefs and higher self-esteem but lower mathematics grades, lower school self-concept beliefs, and lower life satisfaction. The predictive modelling also revealed that at Grade 5 and Grade 9, greater numbers of books at home (which may reflect aspects of socio-cultural advantage/capital) predicted higher mathematics grades, school self-concept beliefs, and self-esteem. These findings reflect some existing inequalities within Germany, where advantaged circumstances have tended to associate with beneficial educational outcomes, and where gender differences in mathematics performance (favouring boys) have tended to increase during secondary school (Autorengruppe Bildungsberichterstattung, 2020, 2022). These (and other) equity and inclusion issues remain relevant to education and also future employment, especially in science-related areas including mathematics (Acatech, 2022; Anger et al., 2022; Autorengruppe Bildungsberichterstattung, 2022; Jeanrenaud, 2020). Further research could continue to explore when and why inequalities (such as across boys and girls) may arise, and also whether and how school contexts and environments are influential, in order to understand what can be done to reduce or eliminate associations between students' backgrounds and their grades and confidence. Ideally, all students should be able to gain high grades, rather than there being differences across genders and numbers of books at home.

4.1. Limitations and areas for future exploration

Any research study can only provide a specific perspective, given particular aims and approaches, and can only provide partial insights into complex processes that occur within wider (and intersecting) aspects of life. Accordingly, the research presented here involved considering calibration bias within predictive modelling, given that this approach had received less focus within prior research. Nevertheless, correlations and predictive modelling can only reveal associations and cannot definitively establish causality. Given the potential for reciprocal associations (where research cannot easily or definitively establish causality), the predictive modelling instead provided multiple perspectives; for example, one model predicted mathematics test calibration bias using mathematics grades, self-concept beliefs, and other predictors, while another model predicted mathematics grades using mathematics test calibration bias, self-concept beliefs, and other predictors. Quantitative research can establish the plausibility of associations, but cannot easily show or establish underlying processes and mechanisms including why and how students think and form beliefs in particular ways; undertaking qualitative research such as interviews could allow students to convey their personal narratives and thought processes to help clarify such matters.

The predictive modelling revealed some areas for future exploration. For example, when predicting mathematics grades, school self-concept, and general life satisfaction at Grade 9 using Grade 9 predictors, being a boy (compared to being a girl) was initially positively associated with the outcomes (when the models only accounted for school and home factors) but was ultimately negatively associated with the outcomes (when the models accounted for school and home factors and also accounted for the students' other views). These changes in predictive coefficients across different arrays of predictors (where coefficients switch between being positive and being negative) may suggest that 'mediation' occurs: the initial appearance of an association between gender and grades (for example) may reflect an underlying gender difference in another aspect of life (for example such as mathematics self-concept beliefs) that then associates with grades. However, further research would need to apply more complex models and consider mediation in more detail in order to gain greater insights.

The results presented here highlighted that, on average, the students tended towards over-confidence. In wider research, students have often, on average, exhibited some over-confidence (for example: Bouffard et al., 1998; Bouffard et al., 2011; Desoete & Roeyers, 2006; Gramzow et al., 2003; Kruger & Dunning, 1999; Winne & Jamieson-Noel, 2002). The research presented here focused on predictive modelling, but future research may also benefit from returning to the traditional approach of considering groups of under-confident, accurate, and over-confident students, but with further reflection on how 'accurate' or 'reasonable' confidence could or should be defined (within the context of many students tending towards over-confidence). Further analytical approaches may also be necessary in order to consider whether smaller extent of over-confidence may be beneficial while larger extents of over-confidence may be detrimental.

Calibration bias (the extent of under-confidence/over-confidence) in mathematics tasks may or may not reflect calibration bias in subject-level mathematics matters or in wider educational matters. Generally, research focusing on task-level calibration bias needs to assume that any task-level processes of self-evaluation reflect someone's wider processes of self-evaluation or metacognitive monitoring; essentially, research needs to assume that someone's generalised tendency towards a confidence bias are reflected within their specific task-level responses (Sheldrake, 2016; Stone, 2000). Students' task-level confidence and subject-level confidence have indeed associated, where confidence can be broadly conceptualised within a hierarchical structure (Bong, 1997, 2002). Nevertheless, someone's confidence on particular tasks may be influenced by their experiences of those particular tasks, while someone's confidence in their subject-level performance may be influenced by more or different factors.

The presented results show associations, but any underlying influence mechanisms and processes remain unclear. Further research would be beneficial in order to consider, more specifically and explicitly, whether/how particular aspects of education (such as teaching/learning approaches) and/or wider aspects of life may influence someone to become under-confident or over-confident. Existing research has suggested that many factors may be relevant to confidence biases, which may potentially (or partially) act as influence mechanisms or processes. These include (for example): experiences when undertaking tasks including perceptions of task difficulty (Boekaerts & Rozendaal, 2010; Chen, 2003; Finn & Tauber, 2015; Rinne & Mazzocco, 2014); comparisons against peers (Dunning et al., 2004; Klar & Giladi, 1999); and wider motivations towards self-enhancement or self-protection (Gramzow, 2010; Jiang & Kleitman, 2015; Sedikides & Alicke, 2012). It is also possible that different students may show under-confidence/over-confidence for different reasons. For example, a study of university students in the United States of America revealed that over-confidence linked with low previous performance predicted lower subsequent performance, while over-confidence linked with a motivation to achieve predicted higher subsequent performance, suggesting that any benefits or detriments of calibration bias may be conditional (Gramzow et al., 2003).

4.2. Conclusions and wider implications

The results presented here highlighted that, on average, students conveyed some over-confidence in their evaluations of their performance on mathematics tasks at Grade 5 and at Grade 9. At Grade 5 and at Grade 9, higher subject-level grades in mathematics predicted lower calibration bias (towards the direction of under-confidence). At Grade 9, higher subject-level mathematics confidence predicted higher calibration bias (towards the direction of over-confidence). These findings affirm the possibility that students' calibration bias (the extent of under-confidence/over-confidence) on particular mathematics tasks may link with their generalised subject-level mathematics performance and beliefs.

At Grade 5, calibration bias negatively predicted mathematics grades, over and above the array of other predictors, but had no predictive association with the other outcomes. At Grade 9, calibration bias positively predicted mathematics self-concept, school self-concept, and self-esteem, over and above the array of other predictors. These differences from Grade 5 to Grade 9 are an important and new insight, which suggests that education may need to increasingly consider under-confidence/over-confidence as students grow older.

Finally, a wider and practical implication is that the measurement and analytical approach from this study can be applied within teaching and learning: asking students to self-evaluate and estimate their expected number of correct answers (or their grade or their confidence for each particular question) is something that can be included within or appended to almost any assessment, relatively simply and easily. Information could then be considered by teachers (as part of adapting teaching for their students, such as giving attention to areas where students may feel confident but may not necessarily give correct answers) and also by students themselves (as part of self-evaluating and regulating their own learning such as determining areas to revise).

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

This paper uses data from the National Educational Panel Study (NEPS): Starting Cohort 3 – 5th Grade, doi:10.5157/NEPS:SC3:9.0.0. From 2008 to 2013, NEPS data were collected as part of the Framework Programme for the Promotion of Empirical Educational Research funded by the German Federal Ministry of Education and Research (BMBF). As of 2014, the NEPS survey is carried out by the Leibniz Institute for Educational Trajectories (LIfBi) at the University of Bamberg in cooperation with a nationwide network.

Supplementary materials

Supplementary material associated with this article can be found, in the online version, at doi:10.1016/j.ijer.2022.102085.

References

Acatech. (2022). *MINT Nachwuchsbarometer 2022*. Munich: Acatech.

Anger, C., Kohlisch, E., Koppel, O., & Plünnecke, A. (2022). *MINT-Frühjahrsreport 2022*. Köln: Institut der deutschen Wirtschaft.

Aßmann, C., Steinhauer, H. W., Kiesl, H., Koch, S., Schönberger, B., Müller-Kuller, A., & Blossfeld, H.-P. (2011). 4 Sampling designs of the National Educational Panel Study: Challenges and solutions. *Zeitschrift für Erziehungswissenschaft, 14*(2 Supplement), 51–65. https://doi.org/10.1007/s11618-011-0181-8

Autorengruppe Bildungsberichterstattung. (2020). *Bildung in Deutschland 2020: Ein indikatorengestützter Bericht mit einer Analyse zu Bildung in einer digitalisierten Welt*. Bielefeld: Wbv Media. https://doi.org/10.3278/6001820gw

Autorengruppe Bildungsberichterstattung. (2022). *Bildung in Deutschland 2022: Ein indikatorengestützter Bericht mit einer Analyse zum Bildungspersonal*. Bielefeld: Wbv Media. https://doi.org/10.3278/6001820hw

Bandura, A. (1977). Self-efficacy: Toward a unifying theory of behavioral change. *Psychological Review, 84*(2), 191–215. https://doi.org/10.1037/0033-295X.84.2.191

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

Blossfeld, H.-P., & Roßbach, H.-G. (2019). Education as a lifelong process: The German National Educational Panel Study (NEPS). *Edition ZfE* (2nd edition). Springer VS.

Education as a lifelong process – The German National Educational Panel Study (NEPS)Blossfeld, H.-P., Roßbach, H.-G., & von Maurice, J. (Eds.), Education as a lifelong process – The German National Educational Panel Study (NEPS). *Zeitschrift für Erziehungswissenschaft, 14*(2 Supplement), (2011). Retrieved from http://link.springer.com/journal/11618/14/2/suppl/page/1.

Blossfeld, H.-P., von Maurice, J., & Schneider, T. (2011). 1 The National Educational Panel Study: Need, main features, and research potential. *Zeitschrift für Erziehungswissenschaft, 14*(2 Supplement), 5–17. https://doi.org/10.1007/s11618-011-0178-3

Boekaerts, M. (1999). Self-regulated learning: Where we are today. *International Journal of Educational Research, 31*(6), 445–457. https://doi.org/10.1016/S0883-0355(99)00014-2

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. https://doi.org/10.1016/j.learninstruc.2009.03.002

Bol, L., Hacker, D., O'Shea, P., & Allen, D. (2005). The influence of overt practice, achievement level, and explanatory style on calibration accuracy and performance. *The Journal of Experimental Education, 73*(4), 269–290. https://doi.org/10.3200/JEXE.73.4.269-290

Bol, L., Hacker, D., Walck, C., & Nunnery, J. (2012). The effects of individual or group guidelines on the calibration accuracy and achievement of high school biology students. *Contemporary Educational Psychology, 37*(4), 280–287. https://doi.org/10.1016/j.cedpsych.2012.02.004

Bong, M. (1997). Generality of academic self-efficacy judgments: Evidence of hierarchical relations. *Journal of Educational Psychology, 89*(4), 696–709. https://doi.org/10.1037/0022-0663.89.4.696

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. https://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. https://doi.org/10.1023/A:1021302408382

Bonneville-Roussy, A., Bouffard, T., & Vezeau, C. (2017). Trajectories of self-evaluation bias in primary and secondary school: Parental antecedents and academic consequences. *Journal of School Psychology, 63*, 1–12. https://doi.org/10.1016/j.jsp.2017.02.002

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. https://doi.org/10.1016/j.ijer.2011.08.001

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. https://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. https://doi.org/10.1111/j.2044-8279.1998.tb01294.x

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. https://doi.org/10.1016/j.ijer.2011.08.003

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

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. https://doi.org/10.1016/j.lindif.2003.08.003

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd edition). Hillsdale, New Jersey: Lawrence Earlbaum Associates.

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

Dufner, M., Denissen, J., van Zalk, M., Matthes, B., Meeus, W., van Aken, M., & Sedikides, C. (2012). Positive intelligence illusions: On the relation between intellectual self-enhancement and psychological adjustment. *Journal of Personality, 80*(3), 537–571. https://doi.org/10.1111/j.1467-6494.2011.00742.x

Dunning, D., Heath, C., & Suls, J. (2004). Flawed self-assessment: Implications for health, education, and the workplace. *Psychological Science in the Public Interest, 5*(3), 69–106. https://doi.org/10.1111/j.1529-1006.2004.00018.x

EACEA [European Commission, European Education and Culture Executive Agency, Eurydice]. (2015). *Mathematics education in Europe: Common challenges and national policies*. Brussels: Publications Office of the European Union. https://doi.org/10.2797/72660

EACEA [European Commission, European Education and Culture Executive Agency, Eurydice]. (2022). *Increasing achievement and motivation in mathematics and science learning in schools*. Luxembourg: Publications Office of the European Union. https://doi.org/10.2797/031821

Efklides, A. (2011). Interactions of metacognition with motivation and affect in self-regulated learning: The MASRL model. *Educational Psychologist, 46*(1), 6–25. https://doi.org/10.1080/00461520.2011.538645
Finn, B., & Tauber, S. (2015). When confidence is not a signal of knowing: How students' experiences and beliefs about processing fluency can lead to miscalibrated confidence. *Educational Psychology Review, 27*(4), 567–586. https://doi.org/10.1007/s10648-015-9313-7
Freund, P. A., & Kasten, N. (2012). How smart do you think you are? A meta-analysis on the validity of self-estimates of cognitive ability. *Psychological Bulletin, 138*(2), 296–321. https://doi.org/10.1037/a0026556
García, T., Rodríguez, C., González-Castro, P., González-Pienda, J. A., & Torrance, M. (2016). Elementary students' metacognitive processes and post-performance calibration on mathematical problem-solving tasks. *Metacognition Learning, 11*(2), 139–170. https://doi.org/10.1007/s11409-015-9139-1
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. https://doi.org/10.1016/j.ijer.2011.08.002
Gramzow, R. (2010). Academic exaggeration: Pushing self-enhancement boundaries. In M. Alicke, & C. Sedikides (Eds.), *Handbook of self-enhancement and self-protection* (pp. 455–471). New York: Guilford Press.
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. https://doi.org/10.1016/S0092-6566(02)00535-4
Guo, J., Nagengast, B., Marsh, H., Kelava, A., Gaspard, H., Brandt, H., & Trautwein, U. (2016). Probing the unique contributions of self-concept, task values, and their interactions using multiple value facets and multiple academic outcomes. *AERA Open, 2*(1), 1–20. https://doi.org/10.1177/2332858415626884
Gustafsson, J.-E., Allodi, M. W., Alin Åkerman, B., Eriksson, C., Eriksson, L., Fischbein, S., & Persson, R. (2010). *School, learning and mental health: A systematic review*. Stockholm: The Royal Academy of Sciences.
Huang, C. (2011). Self-concept and academic achievement: A meta-analysis of longitudinal relations. *Journal of School Psychology, 49*(5), 505–528. https://doi.org/10.1016/j.jsp.2011.07.001
Jansen, M., Scherer, R., & Schroeders, U. (2015). Students' self-concept and self-efficacy in the sciences: Differential relations to antecedents and educational outcomes. *Contemporary Educational Psychology, 41*, 13–24. https://doi.org/10.1016/j.cedpsych.2014.11.002
Jeanrenaud, Y. (2020). *MINT. Warum nicht? Zur Unterrepräsentation von Frauen in MINT, speziell IKT, deren Ursachen, Wirksamkeit bestehender Maßnahmen und Handlungsempfehlungen. Expertise für den Dritten Gleichstellungsbericht der Bundesregierung*. Berlin: Institut für Sozialarbeit und Sozialpädagogik /Bundesministerium für Familie, Senioren, Frauen und Jugend.
Jiang, Y., & Kleitman, S. (2015). Metacognition and motivation: Links between confidence, self-protection and self-enhancement. *Learning and Individual Differences, 37*, 222–230. https://doi.org/10.1016/j.lindif.2014.11.025
Jiang, Y., Song, J., Lee, M., & Bong, M. (2014). Self-efficacy and achievement goals as motivational links between perceived contexts and achievement. *Educational Psychology, 34*(1), 92–117. https://doi.org/10.1080/01443410.2013.863831
Kim, Y.-H., & Chiu, C.-Y. (2011). Emotional costs of inaccurate self-assessments: Both self-effacement and self-enhancement can lead to dejection. *Emotion, 11*(5), 1096–1104. https://doi.org/10.1037/a0025478
Klar, Y., & Giladi, E. (1999). Are most people happier than their peers, or are they just happy? *Personality and Social Psychology Bulletin, 25*(5), 586–595. https://doi.org/10.1177/0146167299025005004
KMK [Kultusminister Konferenz]. (2022a). *Bildungsstandards für das Fach Mathematik Erster Schulabschluss (ESA) und Mittlerer Schulabschluss (MSA)*. Berlin: Sekretariat der Ständigen Konferenz der Kultusminister der Länder in der Bundesrepublik Deutschland.
KMK [Kultusminister Konferenz]. (2022b). *Bildungsstandards für das Fach Mathematik Primarbereich*. Berlin: Sekretariat der Ständigen Konferenz der Kultusminister der Länder in der Bundesrepublik Deutschland.
Kruger, J., & Dunning, D. (1999). Unskilled and unaware of it: How difficulties in recognizing one's own incompetence lead to inflated self-assessments. *Journal of Personality and Social Psychology, 77*(6), 1121–1134. https://doi.org/10.1037/0022-3514.77.6.1121
Lichtenstein, S., & Fischhoff, B. (1977). Do those who know more also know more about how much they know? *Organizational Behavior and Human Performance, 20*(2), 159–183. https://doi.org/10.1016/0030-5073(77)90001-0
Lockl, K. (2013). *Competencies: Assesment of procedural metacognition starting cohorts 2-6 (Version: October 2013)*. Bamberg: University of Bamberg, National Educational Panel Study.
Marsh, H., Kuyper, H., Seaton, M., Parker, P., Morin, A., Möller, J., & Abduljabbar, A. (2014). Dimensional comparison theory: An extension of the internal/external frame of reference effect on academic self-concept formation. *Contemporary Educational Psychology, 39*(4), 326–341. https://doi.org/10.1016/j.cedpsych.2014.08.003
Meixner, S., Schiller, D., von Maurice, J., & Engelhardt-Wölfler, H. (2011). 19 Data protection issues in the National Educational Panel Study. *Zeitschrift für Erziehungswissenschaft, 14*(2 Supplement), 301–313. https://doi.org/10.1007/s11618-011-0191-6
Möller, J., & Marsh, H. (2013). Dimensional comparison theory. *Psychological Review, 120*(3), 544–560. https://doi.org/10.1037/a0032459
Möller, J., Retelsdorf, J., Köller, O., & Marsh, H. (2011). The reciprocal internal/external frame of reference model: An integration of models of relations between academic achievement and self-concept. *American Educational Research Journal, 48*(6), 1315–1346. https://doi.org/10.3102/0002831211419649
Multon, K., Brown, S., & Lent, R. (1991). Relation of self-efficacy beliefs to academic outcomes: A meta-analytic investigation. *Journal of Counseling Psychology, 38*(1), 30–38. https://doi.org/10.1037/0022-0167.38.1.30
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. https://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. https://doi.org/10.1016/j.ijer.2011.08.004
NEPS. (2019). *National Educational Panel Study (NEPS): Starting cohort grade 5 (SC3), Version 9.0.0.* doi:10.5157/NEPS:SC3:9.0.0.
Neumann, I., Duchhardt, C., Grüßing, M., Heinze, A., Knopp, E., & Ehmke, T. (2013). Modeling and assessing mathematical competence over the lifespan. *Journal for Educational Research Online, 5*(2), 80–109.
OECD. (2013). *PISA 2012 assessment and analytical framework: Mathematics, reading, science, problem solving and financial literacy*. OECD Publishing. https://doi.org/10.1787/9789264190511-en
OECD. (2015a). *How's life? 2015: Measuring well-being*. Paris: OECD Publishing. https://doi.org/10.1787/how_life-2015-en
OECD. (2015b). *The ABC of gender equality in education: Aptitude, behaviour, confidence*. Paris: OECD Publishing. https://doi.org/10.1787/9789264229945-en
OECD. (2015c). *Universal basic skills: What countries stand to gain*. Paris: OECD Publishing. https://doi.org/10.1787/9789264234833-en
Phillips, J., & Gully, S. (1997). Role of goal orientation, ability, need for achievement, and locus of control in the self-efficacy and goal-setting process. *Journal of Applied Psychology, 82*(5), 792–802. https://doi.org/10.1037/0021-9010.82.5.792
Pieschl, S. (2009). Metacognitive calibration—An extended conceptualization and potential applications. *Metacognition and Learning, 4*(1), 3–31. https://doi.org/10.1007/s11409-008-9030-4
Regan, E., & DeWitt, J. (2015). Attitudes, interest and factors influencing STEM enrolment behaviour: An overview of relevant literature. In E. K. Henriksen, J. Dillon, & J. Ryder (Eds.), *Understanding student participation and choice in science and technology education* (pp. 63–88). Dordrecht: Springer. https://doi.org/10.1007/978-94-007-7793-4_5.
Rinne, L., & Mazzocco, M. (2014). Knowing right from wrong in mental arithmetic judgments: Calibration of confidence predicts the development of accuracy. *PLoS One, 9*(7), e98663. https://doi.org/10.1371/journal.pone.0098663
Rosenberg, M. (1965). *Society and adolescent self-image*. Princeton, New Jersey: Princeton University.
Rytkönen, K., Aunola, K., & Nurmi, J.-E. (2007). Do parents' causal attributions predict the accuracy and bias in their children's self-concept of maths ability? A longitudinal study. *Educational Psychology, 27*(6), 771–788. https://doi.org/10.1080/01443410701309316
Sáinz, M., & Upadyaya, K. (2016). Accuracy and bias in Spanish secondary school students' self-concept of math ability: The influence of gender and parental educational level. *International Journal of Educational Research, 77*, 26–36. https://doi.org/10.1016/j.ijer.2016.02.009

Schraw, G. (2009). A conceptual analysis of five measures of metacognitive monitoring. *Metacognition and Learning, 4*(1), 33–45. https://doi.org/10.1007/s11409-008-9031-3

Seaton, M., Parker, P., Marsh, H., Craven, R., & Yeung, A. S. (2014). The reciprocal relations between self-concept, motivation and achievement: Juxtaposing academic self-concept and achievement goal orientations for mathematics success. *Educational Psychology: An International Journal of Experimental Educational Psychology, 34*(1), 49–72. https://doi.org/10.1080/01443410.2013.825232

Sedikides, C., & Alicke, M. (2012). Self-enhancement and self-protection motives. In R. Ryan (Ed.), *The Oxford handbook of human motivation* (pp. 303–322). Oxford: Oxford University Press. https://doi.org/10.1093/oxfordhb/9780195399820.013.0017.

Shavelson, R., Hubner, J., & Stanton, G. (1976). Self-concept: Validation of construct interpretations. *Review of Educational Research, 46*(3), 407–441. https://doi.org/10.2307/1170010

Sheldrake, R. (2016). Confidence as motivational expressions of interest, utility, and other influences: Exploring under-confidence and over-confidence in science students at secondary school. *International Journal of Educational Research, 76*, 50–65. https://doi.org/10.1016/j.ijer.2015.12.001

Sheldrake, R., Mujtaba, T., & Reiss, M. J. (2014). 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. *International Journal of Educational Research, 64*, 49–61. https://doi.org/10.1016/j.ijer.2013.10.008

Sheldrake, R., Mujtaba, T., & Reiss, M. J. (2015). Students' intentions to study non-compulsory mathematics: The importance of how good you think you are. *British Educational Research Journal, 41*(3), 462–488. https://doi.org/10.1002/berj.3150

Skaalvik, E., Federici, R., & Klassen, R. (2015). Mathematics achievement and self-efficacy: Relations with motivation for mathematics. *International Journal of Educational Research, 72*, 129–136. https://doi.org/10.1016/j.ijer.2015.06.008

Snijders, T., & Bosker, R. (2012). *Multilevel analysis: An introduction to basic and advanced multilevel modeling* (2nd edition). London: SAGE Publications.

Stone, N. (2000). Exploring the relationship between calibration and self-regulated learning. *Educational Psychology Review, 12*(4), 437–475. https://doi.org/10.1023/A:1009084430926

Suldo, S., Riley, K., & Shaffer, E. (2006). Academic correlates of children and adolescents' life satisfaction. *School Psychology International, 27*(5), 567–582. https://doi.org/10.1177/0143034306073411

Talsma, K., Schüz, B., Schwarzer, R., & Norris, K. (2018). I believe, therefore I achieve (and vice versa): A meta-analytic cross-lagged panel analysis of self-efficacy and academic performance. *Learning and Individual Differences, 61*, 136–150. https://doi.org/10.1016/j.lindif.2017.11.015

Taskinen, P., Schütte, K., & Prenzel, M. (2013). Adolescents' motivation to select an academic science-related career: The role of school factors, individual interest, and science self-concept. *Educational Research and Evaluation, 19*(8), 717–733. https://doi.org/10.1080/13803611.2013.853620

Trautwein, U., Marsh, H., Nagengast, B., Lüdtke, O., Nagy, G., & Jonkmann, K. (2012). Probing for the multiplicative term in modern expectancy–Value theory: A latent interaction modeling study. *Journal of Educational Psychology, 104*(3), 763–777. https://doi.org/10.1037/a0027470

Usher, E., & Pajares, F. (2008a). Self-efficacy for self-regulated learning: A validation study. *Educational and Psychological Measurement, 68*(2), 443–463. https://doi.org/10.1177/0013164407308475

Usher, E., & Pajares, F. (2008b). Sources of self-efficacy in school: Critical review of the literature and future directions. *Review of Educational Research, 78*(4), 751–796. https://doi.org/10.3102/0034654308321456

Weinert, S., Artelt, C., Prenzel, M., Senkbeil, M., Ehmke, T., & Carstensen, C. (2011). 5 Development of competencies across the life span. *Zeitschrift für Erziehungswissenschaft, 14*(2 Supplement), 67–86. https://doi.org/10.1007/s11618-011-0182-7

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

Winne, P., & Jamieson-Noel, D. (2002). Exploring students' calibration of self reports about study tactics and achievement. *Contemporary Educational Psychology, 27*(4), 551–572. https://doi.org/10.1016/S0361-476X(02)00006-1

Zell, E., & Krizan, Z. (2014). Do people have insight into their abilities? A metasynthesis. *Perspectives on Psychological Science, 9*(2), 111–125. https://doi.org/10.1177/1745691613518075

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.

Zimmerman, B., & Moylan, A. (2009). Self-regulation: Where metacognition and motivation intersect. In D. Hacker, J. Dunlosky, & A. Graesser (Eds.), *Handbook of metacognition in education* (pp. 299–315). New York: Routledge.

Zimmerman, B., & Schunk, D. (2011). Self-regulated learning and performance: An introduction and an overview. In B. Zimmerman, & D. Schunk (Eds.), *Handbook of self-regulation of learning and performance* (pp. 1–12). Abingdon: Routledge.