European Journal of Psychology of Education (2024) 39:1537–1559
https://doi.org/10.1007/s10212-024-00804-x

When competence and confidence are at odds: a cross-country examination of the Dunning–Kruger effect

Kajsa Yang Hansen[1] · Cecilia Thorsen[2] · Jelena Radišić[3] · Francisco Peixoto[4] · Anu Laine[5] · Xin Liu[3]

Received: 13 July 2023 / Revised: 21 December 2023 / Accepted: 28 January 2024 /
Published online: 19 February 2024

Abstract
Research has shown that some students who underperform in mathematics overestimate their performance, while others who excel in mathematics underestimate it. Looking at this mismatch of performance and confidence judgement—the Dunning–Kruger effect (DKE)—the current study investigates how well students' confidence judgement and item-specific mathematics competence relate with each other and whether such a relationship differs across six European countries (i.e., Norway, Sweden, Finland, Estonia, Serbia and Portugal). We also examine whether perceived competence, mathematics identity, gender, socioeconomic status and immigration background predict this mismatch and whether these demographic factors function differently between the examined countries. The results show that the DKE could be found across grades three and four in all six countries. However, there are country-specific patterns regarding the relationship between performance, mathematics identity and perceived competence; the DKE; and how different demographic variables predict its occurrences in particular subpopulations.

Keywords Dunning–Kruger effect · Confidence judgement · Perceived competence · Mathematics · SES · Gender · Immigration background

Numerous studies have investigated the association between self-assessed and objectively measured performance, which is commonly called calibration accuracy (Gignac & Zajenkowski, 2020; Wollenschläger et al., 2016). It is widely acknowledged that individuals tend to overestimate or underestimate their performance in various contexts (Coutinho et al., 2021; De Bruin et al., 2017; Dunning et al., 2003; Sanchez & Dunning, 2018) and that confidence judgement is linked to multiple factors (Christopher et al., 2023; Efklides, 2011; Geraci et al., 2023; Serra & DeMarree, 2016; Stankov et al., 2012). However, misalignment between confidence and performance can have long-term consequences for people's lives (Geraci et al., 2023). Within the educational context, students' ability to accurately evaluate their performance may allow them to self-regulate their learning (Andrade, 2019; Hosein & Harle, 2018). Conversely, an inaccurate self-assessment could hamper this process, leading them to underutilise their capabilities (Vonkova et al., 2021) or even make

Extended author information available on the last page of the article suboptimal career choices because of underestimating their abilities (Ehrlinger & Dunning, 2003). For example, an overly critical self-assessment of one's math abilities significantly contributes to girls' lower motivation in STEM subjects and their limited consideration of STEM careers (OECD, 2015).

The Dunning–Kruger effect (DKE) refers to the cognitive bias where individuals with low abilities or knowledge overestimate their skills or expertise, while those with high abilities or knowledge tend to underestimate their skills relative to others (Dunning, 2011, 2015; Kruger & Dunning, 1999). Although the DKE is pervasive across domains (Erickson & Heit, 2015; Ernst et al., 2023; Holling & Preckel, 2005; Kruger & Dunning, 1999; Sheldon et al., 2014), most research has focused on upper-secondary or university students (Christopher et al., 2023; Geraci et al., 2023; Vonkova et al., 2021) and single-country settings (Coutinho et al., 2021; Christopher et al., 2023; Vonkova et al., 2022). Limited attention has been given to younger populations (Boekaerts & Rozendaal, 2010; Dapp & Roebers, 2021) and exploring the confidence–performance dynamics across cultures (Morony et al., 2013; Williams & Williams, 2010) or between similar cultures, that is, across Europe.

The present study examines the relationship between students' mathematics performance and confidence judgements of their item-specific mathematics competence. Next, we investigate cross-country DKE patterns and the influence of demographics, motivational, and identity constructs in connection with the DKE in primary mathematics classrooms.

The role of self-assessment and confidence judgements

Competent decision-making is essential in daily life and any profession (Kleitman et al., 2019), where accuracy and confidence judgement are key ingredients. Accurate assessment of one's performance is crucial for suitable self-regulation, such as in learning processes (Andrade, 2019; Hosein & Harle, 2018) or career decision-making (Ehrlinger & Dunning, 2003). On the contrary, inaccurate self-assessment could hinder these processes, preventing individuals from utilising their capabilities to their full potential (Vonkova et al., 2021).

Andrade (2019) acknowledges a broad spectrum of self-assessment definitions in education that encompass an assessment of one's abilities (Brown & Harris, 2013), processes (Panadero et al., 2016), and products (Epstein et al., 2008). Despite this variety, the definitions work because each object of assessment—competence, process and product—is influenced by self-feedback. At the same time, much of the research on self-assessment is oriented towards the accuracy of students' judgements (Andrade, 2019; Brown & Harris, 2013), yielding mixed results regarding the relationship between student self-ratings and other objective measures ranging from weak (0.20) to strongly positive (0.80).

Another critical component of decision-making is confidence judgement, representing the subjective metacognitive experience stemming from accuracy judgements theory (see Efklides, 2006, 2008, 2011; Stankov, 1999, 2019; Stankov et al., 2012; Stankov et al., 2014). Stankov et al. (2012) suggest that confidence is a robust individual difference situated somewhere between ability and personality (Stankov & Lee, 2008), implementing not only those processes related to task performance but also certainty in beliefs about hypothetical events (Kleitman & Stankov, 2007).

Confidence judgements serve as a metacognitive monitoring measure (Stankov, 2019), and calibration is essential for their effectiveness (Efklides, 2008). Confidence judgements also act as the 'gatekeeper' in the decision-making process (Gilovich et al., 2002), with greater confidence expected to be associated with more competent performance (Kleitman et al., 2019; Stankov et al., 2014). However, individuals can make incorrect judgements of their ability despite high confidence, making overconfidence one of the most significant cognitive biases. Nevertheless, overconfidence can be an adaptive response to certain situations (Parker et al., 2012), warranting a thorough examination of individual differences in confidence and performance and their potential miscalibration (Kleitman et al., 2019).

The Dunning–Kruger effect—to be or not to be?

The Dunning–Kruger effect arises when an individual's relatively low competence prevents them from accurately assessing their performance level (Kruger & Dunning, 1999) because low performers are 'operating from incomplete and corrupted knowledge, they would make many mistakes and not recognise those mistakes as they made them' (Dunning, 2011, p. 260). Conversely, those who excel in a given area tend to underestimate their skills compared with others, thinking the task is simple for everyone. Thus, the DKE can be viewed as the problem of systematic individual differences in meta-cognition (Schlösser et al., 2013).

The manifestation of the DKE varies across different domains (Ernst et al., 2023; Holling & Preckel, 2005) or tasks (Tashiro et al., 2021), but its pervasive influence extends to the cognitive, social and emotional domains (Erickson & Heit, 2015; Kruger & Dunning, 1999; Sheldon et al., 2014). Although capturing the magnitude of the DKE is challenging, a recent meta-analysis reports a mean correlation of 0.29 between self-assessed performance and objectively measured performance (Zell & Krizan, 2014). Some general approaches have been applied to detect the presence of the DKE. One method involves using objective measures of performance or knowledge, such as standardised tests or evaluations by domain experts. By comparing self-assessed competence with these objective measures, it is possible to identify those cases in which people overestimate their abilities or knowledge. However, criticism has also been raised, suggesting that the DKE may be a statistical artefact (Gignac & Zajenkowski, 2020; Magnus & Peresetsky, 2022; Sullivan et al., 2019). Nevertheless, studies refuting DKE can explain only a portion of the effect (see Ehrlinger et al., 2008). Conversely, the evidence supporting the DKE suggests the presence of metacognitive differences in online monitoring between low and high performers (McIntosh et al., 2019). Even so, a call for more robust methodologies in assessing the DKE led to more diversity in methodological approaches and a more robust screening of different DKE correlates (Gignac & Zajenkowski, 2020).

The confidence–competence relationship and its correlates

The confidence–competence relationship has attracted much attention, not only in examining the association itself but also in exploring the factors behind this relationship. Stankov and colleagues have consistently demonstrated that confidence is the best predictor of mathematics performance, explaining a substantial portion of the predictive variance of other self-beliefs, such as self-efficacy, self-concept and perceived competence (Morony et al., 2013; Stankov, 1999, 2019; Stankov et al., 2012; Stankov et al., 2014; Stankov & Lee, 2014, 2017). Confidence accounts for approximately 45% or more of the total variance in mathematics performance (Stankov & Lee, 2014, 2017). In general, students' accuracy in self-assessment is linked to their prior mathematical success and overall (mathematics) confidence (Hosein & Harle, 2018). Komarraju and Nadler (2013) also show that students with high confidence levels believe intelligence is malleable, whereas those with low confidence tend to view intelligence as fixed. For the latter, a fixed mindset may discourage engagement in metacognitive monitoring because of this belief in competence predetermination; thus, possible feedback is disregarded, contributing to further misestimation.

Furthermore, little difference is observed between regions when investigating cross-cultural (in)variance in confidence, and confidence remains the primary predictor of mathematics scores (Morony et al., 2013; Stankov & Lee, 2014). Morony et al. (2013) examine two world regions: Confucian Asia (i.e., Hong Kong, Singapore, South Korea and Taiwan) and Europe (Denmark, The Netherlands, Finland, Latvia and Serbia), finding significant differences in calibration but not the construct of confidence (i.e., the invariance in the meaning of the construct was kept across contexts). Europeans display more overconfidence, probably because of the lower overall mathematics scores of students from Serbia and Latvia, while Confucian Asia shows far more cautious estimations. However, differences between the European countries are also noted—the mean confidence level between European countries ranges from 48.29 in Finland to 64.24 in the Netherlands. In a similar study focused on spatial navigation, Walkowiak et al. (2023) show that cultural clusters of countries tend to be similarly confident; however, when observing Europe alone, Germanic and East European countries show the most overconfidence when compared with the Nordic countries.

In mathematics, compared with other academic subjects, students are prone to show a higher degree of overconfidence (Erickson & Heit, 2015). Task difficulty has been found to significantly affect students' self-assessments (Tashiro et al., 2021), with students more likely to underestimate their competence after engaging in a more difficult task and overestimate own competence following an easier one, independent of performance. Boekaerts and Rozendaal (2010) show that fifth graders might struggle with accurately assessing their performance, and calibration differs depending on the subject (e.g., application problems vs. computation problems). They further argue that repeated underestimation of performance may indicate a lack of confidence, ultimately affecting other aspects of motivation and performance. Conversely, students who overestimate their performance may experience higher short-term motivation. They may also become frustrated by the long-term outcomes of their efforts.

Likewise, Tashiro et al. (2021) find that how students perceive their understanding, compared with their peers, significantly affects their assessment. For example, students who later reported lower understanding than their peers were more likely to downgrade their self-assessed understanding. Similarly, we can postulate that perceived competence (i.e., student's perception of competence over achievement-related tasks; Pekrun, 2006) will affect students' confidence judgements, especially given that, for the latter to be present, one should also hold positive perceptions of competence. Furthermore, mathematics identity (i.e., an individual's sense of self in the mathematics domain, Darragh, 2016) is argued to be a critical factor related to the development of student attainment, affect and involvement in mathematics (Eccles, 2009; Miller & Wang, 2019), providing new insights into why some students may underachieve, underestimate their own competence or disengage from mathematics, despite their prior abilities.

Regarding gender differences, the findings are inconsistent. Although some studies show no difference (Hosein & Harle, 2018; Tashiro et al., 2021), others indicate males consistently overestimate their performance (Boekaerts & Rozendaal, 2010; Marks et al., 2018), including cross-country comparisons (Morony et al., 2013). Among the countries examined in Morony et al.'s (2013) study, minimal gender differences are found only in Singapore, while these are far more prominent in other observed countries. Regarding age, it is argued that older learners tend to display greater consistency in their judgements (Guillory & Blankson, 2017; Nagel & Lindsey, 2018) and that consistency improves with experience (Nagel & Lindsey, 2018). The findings regarding the contribution of feedback have been inconsistent (Tashiro et al., 2021; Thawabieh, 2017). Similarly, socioeconomic status (SES) is often included as a covariate in examining the complex confidence–competence relationship and possible influences on students' self-assessment and performance in mathematics. Stankov and Lee (2014, 2017) show that SES accounts for around 30% of the variance explaining performance.

Current study

Previous research has underscored the importance of students' accurate self-assessment of their performance, particularly because of its crucial role in the self-regulation of learning (Andrade, 2019; Hosein & Harle, 2018). Moreover, in other domains of life, competent decision-making holds substantial importance, affecting different life choices (Kleitman et al., 2019).

From an early stage, children's self-assessments reflect their objective performance (Dapp & Roebers, 2021). However, multiple factors affect the process, namely, confidence (e.g., Morony et al., 2013; Stankov, 2019; Stankov et al., 2012), task features (Tashiro et al., 2021), gender (Boekaerts & Rozendaal, 2010; Marks et al., 2018), SES (Stankov & Lee, 2014, 2017) and motivation (Boekaerts & Rozendaal, 2010). These factors can hinder students' capacity to calibrate their judgements. Unfortunately, a cross-cultural examination of this process related to mathematics remains limited (Morony et al., 2013).

The current study focuses on primary students in six European countries, exploring the following research questions:

(1) What is the relationship between students' mathematics performance and confidence judgements of their item-specific mathematics competence? That is, can the DKE be established in younger populations? Given the prevalence of studies in older age groups and university populations (Christopher et al., 2023; Geraci et al., 2023; Vonkova et al., 2021), we expect the phenomenon to be somewhat visible across the observed population
(2) Are there cross-country differences in the DKE? Based on cross-cultural studies on confidence-performance dynamics (Morony et al., 2013), we expect DKE to be detected across all the examined European countries. At the same time, similar to Morony et al. (2013), Stankov and Lee (2014) and Walkowiak et al. (2023), we postulate country-specific patterns will be detected in the younger age groups
(3) Which demographic factors contribute to DKE? Based on previous studies on the effect of gender (Boekaerts & Rozendaal, 2010; Marks et al., 2018) and SES's contribution to explaining performance (Stankov & Lee, 2014, 2017), we examine their contribution in connection to the DKE
(4) How does the DKE align with one's mathematics identity and perceived competence in mathematics? This research question is more exploratory. However, based on some prior studies observing these concepts and mutual relationships (e.g., Darragh, 2016;

Eccles, 2009; Pekrun, 2006), we expect students who exhibit higher levels of perceived competence and mathematics identity to be less susceptible to the DKE

Methods

The current study used the first-wave data collected within international longitudinal research focused on the development of mathematics motivation in primary education—Co-Constructing Mathematics Motivation in Primary Education–A Longitudinal Study in Six European Countries (MATHMot for short)—funded by the Research Council of Norway (grant number 301033). All students in the six participating countries, namely, Estonia, Finland, Norway, Portugal, Serbia and Sweden, that were involved in the first wave were included in the investigation. The distribution of participants in each country and grade is presented in Table 1. The sample size ranged from 843 in Estonia to 1093 in Norway in grade 3 and from 851 in Estonia to 1255 in Portugal in grade 4. In total, 6073 third graders and 5696 fourth graders were included in the analysis.

Variables

In the first wave of the MATHMot data collection, students in both grades 3 and 4 were asked to answer questions related to, among other things, their background (i.e., number of books at home and language spoken at home), their motivation in learning mathematics (i.e., including perceived competence in mathematics), confidence judgement and mathematics identity. Table 2 shows the variables used in the current study and the variable properties of each grade and country.

Perceived competence in mathematics (PCM) was measured using the frequency of students endorsing the five statements about self-perception of their general mathematics competence, such as ‘Math is easy for me’. PCM was delivered as a subscale of the Expectancy Value Scale (Peixoto et al., 2023), which is grounded on Eccles’ situated expectancy value theory (Eccles & Wigfield, 2020). Each statement had four response alternatives, ranging from ‘very often’ to ‘never’. The same items were used for both grades. Cronbach’s alphas for PCM were satisfactory across grades and countries, with a trend of higher values in grade 4 (see Table 2 for details and Appendix).

The construct of students’ mathematics identity (MI) was measured by six statements related to their perceived personal (e.g., I think I am a math person) or recognised perceptions (e.g., My family thinks of me as a math person). The MI scale was adapted and

Table 1 Number of students in grade 3 and grade 4 in all six countries

Country	Grade 3	Grade 4
Estonia	843	851
Finland	896	871
Norway	1093	1042
Portugal	855	1255
Serbia	1021	1138
Sweden	965	939
Total	6073	5696

Table 2 Information on the variables involved in the analyses

Variable	Variable information	Reliability information
Sex	Dummy variable: 1 = boys, 0 = girls	/
Language at home (D_lang)	Dummy variable: 0 = non-native, 1 = native	/
Number of books at home	Categorical variable: 1 = 0–10 books, 2 = 11–25 books, 3 = 26 to 100 books, 4 = 101 to 200 books, 5 ≥ 200 books	/
Perceived competence in mathematics (PCM)	Example item: Math is easy for me	0.778 to 0.858
Mathematics identity	Example item: I think I am a math person	0.744 to 0.833
Mathematics test score	12 test items in grade 3 and 14 items in grade 4	IRT estimated score
Confidence judgement	12 (14) items of confidence judgement corresponding to the math test items. Principle component factor score estimate with pooled mean of six countries 0 and standard deviation 1	/

Cronbach's alpha for PCM in grade 3: Estonia: α = 0.814, Finland: α = 0.809, Norway: α = 0.788, Portugal: α = 0.778, Serbia: α = 0.783 and Sweden: α = 0.797. Grade 4: Estonia: α = 0.858, Finland: α = 0.82, Norway: α = 0.802, Portugal: α = 0.827, Serbia: α = 0.818 and Sweden: α = 0.821. Cronbach's alpha for MI in grade 3: Estonia: α = 0.792, Finland: α = 0.810, Norway: α = 0.788, Portugal: α = 0.744, Serbia: α = 0.752 and Sweden: α = 0.764. Grade 4: Estonia: α = 0.832, Finland: α = 0.816, Norway: α = 0.802, Portugal: α = 0.808, Serbia: α = 0.780 and Sweden: α = 0.80

further advanced based on the prior work of Vincent-Ruz et al. (2018) and Miller and Wang (2019). Each of the six items has four response alternatives, ranging from 'very often' to 'never'. The Cronbach's alpha for both grades in almost all countries was over 0.75. Generally, grade 4 samples observed higher reliability in this construct, ranging from 0.780 in Serbia to 0.832 in Estonia. For the grade 3 samples, Finland had the highest reliability at 0.810 and Portugal had the lowest at 0.744. Earlier analyses on the validation of the EVS scale (Peixoto et al., 2023) have shown a correlation of 0.55 between the PCM and MI measures.

Each student was also given a mathematics test with 12 items in grade 3 and 14 items in grade 4. The test was timed (i.e., 25 min in grade 3 and 30 min in grade 4). The tasks were chosen from the previous TIMSS assessment (Approval IEA-22–022), depicting the curriculum span of the involved countries. The students received one point for each correct answer, resulting in a 12-point maximum score in grade 3 and 14 in grade 4. Math performance scores were estimated using the Rasch model based on all item responses. The item pool included seven shared items in both grades as linking items. The students' estimated math scores were rescaled in the present study within each grade, with an average score of 0 and standard deviation of 1. After each mathematics test item, students were asked how confident they were in solving a task (Kleitman & Stankov, 2007), responding on a 4-point scale (ranging from 'not at all confident' to 'very confident'). These formed the construct of confidence judgement. The construct in the present study captures item-specific self-confidence in mathematics.

The dummy variable language used at home differentiates the native vs. non-native students, indicating the student's immigrant background. Students' sex was also a dummy variable, while the number of books at home, which is a proxy for SES, is 5-point categorical variable (see Table 2).

Table 3 presents the percentage of students in each variable category across countries and grades. The distribution of boys and girls was similar for all subgroups (i.e., grades and countries). Ethnicity, as measured by the frequency of speaking the test language at home, captured students' cultural diversity. With that in mind, Norway had the largest proportion of language diversity, while Serbia had the smallest. For all countries, the largest proportion of students reported having about 26 to 100 books in their homes, ranging from 34% in Portugal to about 40% in Estonia.

Analytical approach

Several steps were taken to investigate the four research questions. First, we tested the structure of the PCM and MI measures. Unidimensionality for each scale was established. Second, the factor scores were calculated for the item-specific confidence judgement, PCM and MI. The factor score was estimated using the pooled data of all countries for each grade. The estimated math score for the three constructs was based on all six countries' pooled data, with an overall mean of 0 and a standard deviation of 1. In doing so, the country-specific mean and standard deviation were relative to a common scale and support cross-country comparisons. In the next step, the mean values of the factor scores of mathematics test scores and item-specific confidence judgement at each quintile of mathematics performance were plotted to study how the two variables posited with each other. Finally, the correlation between mathematics performance scores and confidence judgements was estimated for each quintile. The proportion of males, native individuals, SES (indicated by the number of books at home), PCM and MI was used as controls in both grades and countries to depict the characteristics of individuals in each mathematics quintile.

Table 3 Distribution of variables in each grade and country (values are given in %)

	Variable	Categories	Estonia	Finland	Norway	Portugal	Serbia	Sweden
Grade3	Books	0–10 books	9.1	5.6	10	14.2	6	6.8
		11–25 books	25.6	17.1	16.2	28.2	20.9	20.2
		26–100 books	40.6	42.8	37	33.5	36.2	39.8
		101–200 books	15.2	21.1	18.3	13.6	18.8	19.9
		More than 200 books	9.6	13.4	18.5	10.4	18	13.4
	Sex	Girls	50.7	51.6	49.7	48.7	50.1	54.1
		Boys	49.3	48.4	50.3	51.3	49.9	45.9
	D_lang	Native	79.5	74	60.7	78.8	90.2	74.9
		Non-native	20.5	26	39.3	21.2	9.8	25.1
Grade4	Books	0–10 books	9.3	5.8	8.8	12.4	6.7	8
		11–25 books	22.2	16.2	17.6	28	19.2	21.6
		26–100 books	40.6	37.8	35.9	33.7	38	35.5
		101–200 books	16.8	25.6	21.5	15.3	20.4	18.4
		more than 200 books	11.2	14.5	16.3	10.7	15.7	16.4
	Sex	Girls	53.3	49.4	46.1	50.8	49.5	54.6
		Boys	46.7	50.6	53.9	49.2	50.5	45.4
	D_lang	Native	86.5	73.3	63.1	81.5	91.6	75.5
		Non-native	13.5	26.7	36.9	18.5	8.4	24.5

D_lang, dummy coded language used at home, with 0 = non-native and 1 = native

Results

We start this section by observing the patterns between students' mathematics performance (i.e., total math score), confidence judgement of their item-specific mathematics competence, MI and PCM. The section further examines the relationship between students' mathematics performance and confidence judgement of their item-specific mathematics competence, that is, the DKE effect. This was done by plotting the mean values of the two construct measures on the mathematics quintile scale for each country and grade. The bivariate correlation coefficient and partial correlation of the two constructs were also presented for each mathematics performance quintile. The results section continues by examining cross-country DKE differences in the six European countries and those factors contributing to the more substantial DKE effect.

Comparison of mathematics performance, confidence judgement, math identity and perceived competence

As shown in Appendix, Swedish third-grade children achieved the highest math performance scores (0.30) even though they exhibited the lowest confidence judgement (− 0.32) and a relatively low level of MI (− 0.11). A similar pattern was observed among Estonian third graders, who achieved math performance of 0.29 and low levels of confidence judgement (− 0.16) and MI (− 0.25). In contrast, Norwegian students had the lowest math performance (− 0.28), the lowest MI (− 0.41) and relatively low confidence judgement (− 0.14). Portuguese pupils had the highest confidence judgements (0.46) but only achieved an average level on the mathematics test (− 0.06). The third graders in Serbia achieved relatively low scores in mathematics (− 0.20), and notably, they had the highest level of MI (0.46) and relatively high confidence judgement (0.16). Finnish children achieved at or below average levels in mathematics, but interestingly, they had the highest perceived mathematics competence (0.30). Conversely, their Portuguese counterparts had the lowest PCM (− 0.20), despite having a very high item-specific confidence judgement and MI (0.17).

In grade 4 (for details, see Appendix), Estonian students achieved the highest performance in mathematics (0.37), while their confidence judgement was the lowest (− 0.27) and MI was also relatively low (− 0.21). Norwegian students held the lowest math performance (− 0.17) coupled with low MI scores (− 0.31) and rather low confidence judgements (− 0.15) and PCM (− 0.12). Again, the Portuguese students had the highest confidence judgement (0.48) and above average scores on MI (0.13). Regarding PCM, we have found a similar pattern as the one in grade 3—Finland holds the highest value (0.40) and Portugal the lowest (− 0.21). Fourth graders in Serbia achieved around the average in mathematics and maintained an average PCM. However, their confidence judgement and MI were rather high (0.40 and 0.22, respectively).

Can we observe the Dunning–Kruger effect?

A common pattern is exhibited in Fig. 1, plotting the quintile mean students' mathematics scores and their confidence judgement in grade 3. In all countries, low achievers tended to overestimate their test item-specific confidence judgement compared with their high achiever counterparts. Generally, students who achieved below the third quintile (around the 60th percentile) tended to overestimate their item-specific math competence. However, different country-specific patterns can also be observed when

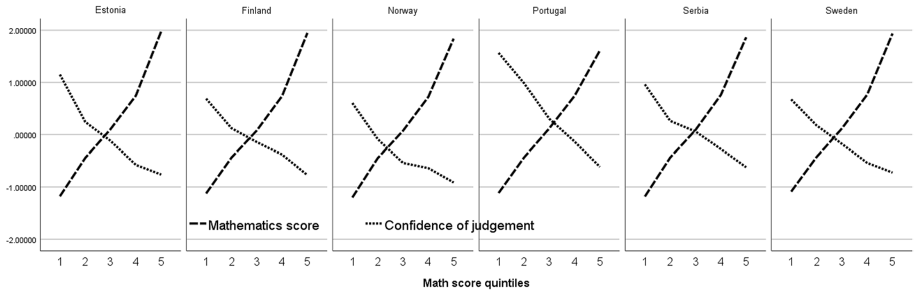

Fig. 1 The mean plot of the mathematics performance score and confidence judgement across each quintile of mathematical competence in grade 3

scrutinising the within-quintile correlation between the mathematics score and confidence judgement (see Table 4).

The analysis revealed a significant negative correlation between students' confidence judgement of their item-specific mathematics competence and their actual mathematics performance across all countries in the total sample, with correlations ranging from − 0.60 in Portugal to − 0.48 in Norway. However, the strength of this negative correlation varied when examining subpopulations of students in each quintile of mathematics performance levels. Specifically, those students in the lowest quintile of mathematics performance in Estonia, Norway, Serbia and Sweden significantly misestimated their item-specific competence, with the highest negative correlation observed in Sweden (− 0.58). The incorrect estimation of students' item-specific math competence was prevalent only in the low-achieving quintile. As the mathematics performance score increased, only one or two countries observed misestimation. In Norway, students in the 4th quintile misestimated their math competence by − 0.23.

Sweden presented a unique case in which students in almost all quintiles significantly wrongly estimated their item-specific math competence, except for those in the 4th quintile.

Table 4 Pearson's correlation coefficients between mathematics performance and confidence judgement for the total sample and within each math score quintile in grade 3

Country		1st quintile	2nd quintile	3rd quintile	4th quintile	5th quintile	Total
Estonia	*r*	− 0.29*	− 0.18	− 0.08	− 0.11	0.06	− 0.55**
	N	75	110	169	172	131	657
Finland	*r*	0.01	− 0.14	0.02	− 0.02	− 0.20	− 0.48**
	N	104	121	147	126	89	587
Norway	*r*	− 0.23**	− 0.06	0.05	− 0.23*	− 0.10	− 0.50**
	N	175	140	122	87	58	582
Portugal	*r*	− 0.19	− 0.09	− 0.28**	− 0.19*	− 0.14	− 0.60**
	N	83	152	192	162	42	631
Serbia	*r*	− 0.28**	− 0.12	− 0.08	− 0.08	− 0.11	− 0.51**
	N	124	182	159	115	68	648
Sweden	*r*	− 0.53**	− 0.26*	− 0.20*	− 0.11	− 0.24**	− 0.49**
	N	37	66	116	125	147	491

**significant at $p < 0.001$ level; *significant $p < 0.05$ level. *r* is Pearson's correlation coefficient

The only significant correlation between mathematics performance and confidence judgement in the 5th quintile was observed in Sweden (−0.24). Conversely, in Finland, the students demonstrated independence in confidence judgements and item-specific mathematics competence, with no significant correlations observed in any of the quintiles. Finally, Portuguese students in the 3rd and 4th quintiles significantly underestimated their math performance (−0.28 and −0.19, respectively).

Figure 2 displays the plotted quintile means of the mathematics scores and students' confidence judgements in grade 4. The same pattern observed in grade 3 was also evident in grade 4, whereby all countries exhibited a misalignment between students' actual mathematics performance and their perceived level of item-specific math competence (see Fig. 2).

Specifically, students with lower mathematics performance levels tended to overestimate their competence. The opposite was also observed; that is, students with higher mathematics scores vastly underestimated their item-specific competence. This trend indicates that individuals with extremely low or high levels of competence may have limited self-awareness of their actual abilities.

As displayed in the 'Total' column in Table 5, a highly significant negative correlation was observed across all countries, ranging from −0.68 in Portugal to −0.52 in Sweden. However, when scrutinised within each quintile, the relationship revealed a different pattern based on the students' mathematics performance. Except for Sweden, all countries in the lowest quintile observed a significant negative correlation between students' mathematics performance and their item-specific confidence judgements, with Portugal exhibiting the strongest correlation (−0.40) and Norway the weakest (−0.18). For the top quintile achievers, significant negative relationships were observed for Estonia (−0.22), Norway (−0.21), Portugal (−0.30) and Serbia (−0.28). In Sweden, the 4th graders appeared to hold more accurate confidence judgements of their item-specific math competence independent of their mathematics performance, and the only mismatch observed was the students in the fourth quintile, who significantly underestimated their math performance (−0.23). Conversely, in Serbia and Portugal, all or almost all their 4th graders demonstrated a lack of appropriate (meta)reflection of their abilities, either underestimating or overestimating their item-specific mathematical competence. All countries in both grades exhibited the Dunning–Kruger effect concerning students' mathematical performance and item-specific confidence judgements. Nonetheless, there were variations in the patterns and severity of this effect. Specifically, fourth graders appeared to be more prone to a lack of (meta)reflection on their competence than third graders. We have found the

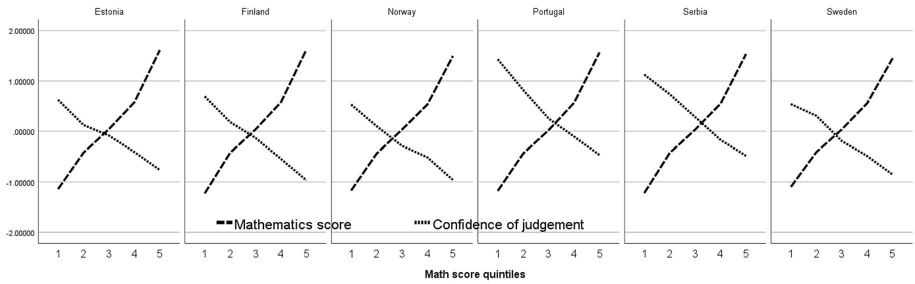

Fig. 2 The mean plot of the mathematics performance score and confidence judgement across each quintile of mathematical competence in grade 4

Table 5 Pearson's correlation coefficients between mathematics performance and confidence judgements for the total sample and within each quintile of math scores in grade 4

Country		1st quintile	2nd quintile	3rd quintile	4th quintile	5th quintile	Total
Estonia	r	−0.37**	−0.08	−0.15	−0.24**	−0.22**	−0.53**
	N	68	89	138	173	197	665
Finland	r	−0.21*	0.09	−0.10	−0.03	−0.16	−0.59**
	N	130	110	124	136	138	638
Norway	r	−0.18*	−0.12	−0.10	−0.08	−0.21*	−0.56**
	N	190	120	139	132	114	695
Portugal	r	−0.40**	−0.14	−0.16*	−0.19*	−0.30**	−0.68**
	N	229	166	171	165	149	880
Serbia	r	−0.32**	−0.32**	−0.24**	−0.25**	−0.28**	−0.62**
	N	109	91	129	150	132	611
Sweden	r	−0.00	−0.10	−0.11	−0.23**	−0.10	−0.52**
	N	76	65	109	149	146	545

** significant at $p < 0.001$ level; * significant $p < 0.05$ level. r is Pearson's correlation coefficient

most considerable disparity in the within-quintile relationship between the two grades in Sweden, Serbia and Portugal.

Who is more susceptible to the Dunning–Kruger effect?

Given that the DKE primarily impacted individuals who lacked expertise in mathematics (i.e., the lowest quintile of math score) but attempted to assess their mathematics competence, we investigated whether additional characteristics of these individuals are more susceptible to the DKE. Table 6 presents the mean differences in the proportions of boys, native students, average SES, PCM and MI across the lowest quintile of math scores in both grades.

As shown in Table 4 and 5, students in the lowest quintile of mathematics performance generally misestimate their item-specific mathematics competence. This subgroup of students comprised a significantly higher number of girls with a migration background, fewer books at home, lowest PCM and low MI compared with their counterparts in the 5th quintile (see supplementary material, Tables A and B for detailed pairwise comparisons). However, particular patterns were also observed. For example, the family migration background did not differ across all quintiles for either third or fourth graders in Estonia and Serbia, countries with the least number of non-native participants. The same was observed for third graders in Portugal. Additionally, no differences between boys and girls were observed in Norway and Serbia in grade 3 and Finland and Sweden in grade 4.

Discussion

The present paper has shed light on the relationship between students' mathematics performance and confidence judgements of their item-specific mathematics competence among primary school students in six European countries. We have also explored demographic factors contributing to the DKE and how the effect aligns with one's MI and PCM.

Table 6 Mean differences in the sociodemographic variables and perceived competence in the lowest quintile of math performance in both grades

	Variables	Estonia		Finland		Norway		Portugal		Serbia		Sweden	
		Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SD
Grade 3	Sex	0.39	0.49	0.39	0.49	0.46	0.50	0.44	0.50	0.48	0.50	0.36	0.48
	D_lang	0.74	0.44	0.61	0.49	0.56	0.50	0.75	0.44	0.90	0.30	0.70	0.46
	Books	2.63	1.10	2.85	0.97	2.89	1.26	2.54	1.21	2.89	1.23	2.82	1.22
	PC	−0.54	0.92	−0.13	0.91	−0.38	1.03	−0.56	0.99	−0.26	0.98	−0.53	1.09
	MI	−0.13	0.93	−0.19	1.02	−0.06	0.98	−0.31	1.03	−0.33	0.95	−0.25	0.88
Grade 4	Sex	0.36	0.48	0.48	0.50	0.44	0.50	0.43	0.50	0.43	0.50	0.38	0.49
	D_lang	0.78	0.41	0.60	0.49	0.55	0.50	0.76	0.43	0.91	0.28	0.67	0.47
	Books	2.62	1.16	2.76	1.08	2.83	1.20	2.49	1.14	2.94	1.19	2.74	1.16
	PC	−0.71	0.89	0.01	0.87	−0.58	0.99	−0.65	0.98	−0.54	0.99	−0.44	1.01
	MI	−0.41	0.81	−0.32	0.99	−0.15	0.95	−0.22	0.93	−0.49	1.02	−0.28	0.97

D_lang, dummy coded language used at home, with 0 = non-native and 1 = native; *PC*, perceived competence in mathematics; *MI*, mathematics identity

Our results showed that overall country patterns remain similar between grades 3 and 4. For example, Norway scored the lowest consistently across math, confidence judgements and MI. Similarly, Portugal (low) and Finland (high) maintained opposite values on PCM and confidence judgements in both grades. These results were consistent with earlier studies on country-specific patterns (Morony et al., 2013; Stankov & Lee, 2014), but they also show the importance of cultural practises and traditions. For example, student self-evaluation has formed a central part of the evaluation in Finnish basic education in recent decades (Pitkänen, 2023), and an earlier study shows that the effect of mathematical competence on self-efficacy is one of the largest in Finland but not vice versa (Williams & Williams, 2010). This supports the idea that experience can improve consistency (Nagel & Lindsey, 2018) and that particular cultural practises have contributed to better calibration among students in Finland. In the case of Serbia, we could easily argue that older learners do not display greater consistency in their judgements, contrary to Guillory and Blankson (2017) and Nagel and Lindsey (2018), because overconfidence is visible in both grades. However, such overconfidence at the national level can also be viewed as a certain cushion, protecting low-achieving individuals (or groups) from experiencing too much stress (Morony et al., 2013).

The second main result of our study is that the DKE was found among pupils in grades 3 and 4 in all studied countries. Low achievers tended to overestimate their test item-specific confidence judgement compared with their high achiever counterparts (Dunning, 2011; Kruger & Dunning, 1999). Thus, the effect is more visible among students scoring below the third quintile (around the 60th percentile). Specifically, students in the lowest quintile of mathematics performance in Estonia, Norway, Serbia and Sweden significantly overestimated their item-specific mathematics competence. These results also confirmed that the effect cannot be merely argued to be a statistical artefact (McIntosh et al., 2019) and that an examination of particular student populations may provide more rigour in examining the DKE and explaining some of its underlying mechanisms.

The results regarding Sweden may be an interesting illustration of that claim. The third-grade students in Sweden significantly overestimated and underestimated their performance. The first and second-quintile students significantly overestimated their performance, and the third and fifth quintiles significantly underestimated their results. One possible explanation for these results might be that the students in grade 3 took the national mathematics tests before the survey's administration. Hence, it might be that the feedback from the national tests has affected the students' assessments of their own competence. Even though the national tests in grade 3 are not high stakes, the students perceived them as such (e.g., Bagger, 2016). Hence, it is possible that the higher-achieving students perceived the test feedback as a sign of failure, lowering their performance estimations (Hosein & Harle, 2018). However, this does not seem to explain the overestimation of performance for lower-achieving students. It is possible that this group is not as affected by the feedback from previous testing because of not being as engaged in metacognitive monitoring compared with their higher-achieving peers (Komarraju & Nadler, 2013). At the same time, Swedish students in grade 4 were much more accurate in their assessment of their own performance. A possible explanation for this grade difference might be found in the assessment environment. In grade 4, grading criteria are introduced, and a stronger focus on grading and summative assessments is embedded in teaching. The introduction of grading into the classroom might result in increased social comparison, affecting how students perceive their competence (Tashiro et al., 2021).

Finally, we observed the impact of demographic factors and perceived competence and MI in connection to the DKE, particularly students who lack competence in mathematics because they are more prone to the DKE. The subgroup is characterised by a significant number of girls from the non-native group, fewer books at home and the lowest PCM compared with their counterparts in the 5th quintile. These results confirm earlier findings favouring boys (Boekaerts & Rozendaal, 2010; Marks et al., 2018), even across countries (Morony et al., 2013). However, because there were no differences between boys and girls observed in Norway and Serbia in grade 3 and Finland and Sweden in grade 4, this provided support for the inconsistency of earlier findings (Hosein & Harle, 2018; Tashiro et al., 2021), again stressing the need for a more nuanced approach in observing DKE correlates and mechanisms (Gignac & Zajenkowski, 2020). Because our data do not allow for differentiation of student background and, for example, the population in Serbia is more compact in connection to the native/non-native divide, warrants further investigation on the contribution of demographic variables.

Finally, MI largely showed a linear increase in values between the 1st to 5th quintiles, similar to PCM. Although the linear increase is not perfect, showing 5th-quintile students to be more conservative makes room for concluding that perceived competence is linked to student estimations and their self-image in mathematics (Miller & Wang, 2019). Earlier, Boekaerts and Rozendaal (2010) have suggested that the importance of student calibrations is not only for confidence judgements but also for long-term motivation, of which perceived competence is a part. In addition, when students grasp what they can and cannot do correctly, it ultimately contributes to their self-image and more robust knowledge of which of their skills need additional improvement and where their strengths lie. All of these are essential in the context of successful self-regulation and learning at school (Andrade, 2019; Hosein & Harle, 2018).

Limitations and further research

Although the data used stemmed from a longitudinal project, the study itself was based on a single wave of data, limiting the ability to observe how the effect evolved over time and establish causal relationships. The nature of the current investigation was more exploratory and descriptive, focusing on detecting the DKE in younger samples and possible differences when observing several European countries simultaneously. Given that the effect can be influenced by various factors—cognitive biases, metacognitive processes and socio-emotional factors (e.g., anxiety, Morony et al., 2013; Stankov & Lee, 2017)—or variability in past performance (Geraci et al., 2023), further investigation into their interplay and specific contributions to the effect will be further explored with the second wave data now being collected. This will allow for a more nuanced exploration of the models examined now and possibly aid in developing a calibration measure between students' confidence judgements and performance as an indicator of the DKE.

Conclusion

The DKE was observed in grades 3 and 4 across all countries in the present study, indicating a general susceptibility to this phenomenon. However, the miscalibration between students' confidence judgement of their performance and their actual mathematics performance appears primarily associated with students with limited mathematics competence. This group of students—a large proportion of girls—can be characterised as children with disadvantaged sociodemographic backgrounds, low self-perception and low identity in mathematics. However, variations in the DKE were also noted across countries, suggesting that different mechanisms and factors may contribute to the miscalibration of competence and confidence judgements within different performance groups or countries. The current study contributes to Dunning–Kruger's research by comparing younger individuals with those within Europe. The findings highlight the importance of educators supporting students, particularly those with weaker skills, in striking a balance between self-confidence and actual competence. This balanced approach can help students identify areas for improvement (Boekaerts & Rozendaal, 2010), ultimately fostering more accurate perceptions and increased confidence in their abilities over time, hence supporting learning (Andrade, 2019; Hosein & Harle, 2018).

Appendix

Table 7 Mean and standard deviation of the variables in each grade and country

Country		Grade 3		Grade 4	
		Mean	SD	Mean	SD
Estonia	Mathematics performance	0.29	1.04	0.37	0.96
	Confidence judgement	−0.16	0.97	−0.27	0.85
	Perceived competence in mathematics	−0.10	0.98	−0.03	1.00
	Student's mathematics identity	−0.25	0.99	−0.21	1.01
Finland	Mathematics performance	0.04	0.99	0.01	1.03
	Confidence judgement	−0.09	0.91	−0.18	.98
	Perceived competence in mathematics	0.30	0.92	0.40	0.87
	Student's mathematics identity	0.10	0.95	0.10	0.92
Norway	Mathematics performance	−0.28	0.98	−0.17	0.98
	Confidence judgement	−0.14	1.04	−0.15	0.94
	Perceived competence in mathematics	−0.09	1.00	−0.12	0.99
	Student's mathematics identity	−0.41	1.05	−0.31	1.04
Portugal	Mathematics performance	−0.06	.85	−0.15	0.98
	Confidence judgement	0.46	1.07	0.48	1.04
	Perceived competence in mathematics	−0.20	1.05	−0.21	1.04
	Student's mathematics identity	0.17	0.91	0.13	0.95
Serbia	Mathematics performance	−0.20	0.93	−0.05	1.02
	Confidence judgement	0.16	0.92	0.22	0.96
	Perceived competence in mathematics	0.06	1.00	0.00	1.01
	Student's mathematics identity	0.46	0.87	0.40	0.92
Sweden	Mathematics performance	0.30	1.05	0.10	0.94
	Confidence judgement	−0.32	0.85	−0.29	0.91
	Perceived competence in mathematics	0.03	0.97	0.06	0.94
	Student's mathematics identity	−0.11	0.94	−0.14	0.97

Table7

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1007/s10212-024-00804-x.

Funding Open access funding provided by University of Gothenburg. Research Council of Norway within FINNUT Programme for Research and Innovation in the Educational Sector (grant number 301033).

Declarations

Conflict of interest The authors declare that they have no conflict of interest.

References

Andrade, H. L. (2019). A critical review of research on student self-assessment. *Frontiers in Education, 4*(87). https://doi.org/10.3389/feduc.2019.00087

Boekaerts, M., & Rozendaal, J. S. (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

Brown, G. T., & Harris, L. R. (2013). Student self-assessment. In J. H. McMillan (Ed.), *SAGE handbook of research on classroom assessment* (pp. 367–393). Sage. https://doi.org/10.4135/9781452218649.n21

Christopher, K. I., Padmakumari, P., & Herbert, H. S. (2023). Presence or absence of Dunning-Kruger effect: Differences in narcissism, general self-efficacy and decision-making styles in young adults. *Current Psychology, 42*, 1444–1455. https://doi.org/10.1007/s12144-021-01461-9

Coutinho, M.V.C., Thomas, J., Alsuwaidi, A. S. M., & Couchman, J.J., (2021). Dunning-Kruger effect: Intuitive errors predict overconfidence on the cognitive reflection test. *Frontiers in Psychology, 12*, https://doi.org/10.3389/fpsyg.2021.603225

Dapp, L. C., & Roebers, C. M. (2021). Metacognition and self-concept: Elaborating on a construct relation in first-grade children. *PLoS One, 16*(4), e0250845. https://doi.org/10.1371/journal.pone.0250845

Darragh, L. (2016). Identity research in mathematics education. *Educational Studies in Mathematics, 93*(1), 19–33. https://doi.org/10.1007/s10649-016-9696-5

De Bruin, A. B. H., Kok, E., Lobbestael, J., & De Grip, A. (2017). The impact of an online tool for monitoring and regulating learning at university: Overconfidence, learning strategy, and personality. *Metacognition Learning, 12*, 21–43. https://doi.org/10.1007/s11409-016-9159-5

Dunning, D. (2011). The Dunning-Kruger effect: On being ignorant of one's own ignorance. In J. Olson & M. P. Zanna (Eds.), *Advances in experimental social psychology* (Vol. 44, pp. 247–296). Elsevier.

Dunning, D. (2015). On identifying human capital: Flawed knowledge leads to faulty judgementss of expertise by individuals and groups. In S. R. Thye & E. J. Lawler (Eds.), *Advances in group processes* (pp. 149–176). Emerald Group Publishing.

Dunning, D., Johnson, K., Ehrlinger, J., & Kruger, J. (2003). Why people fail to recognise their own incompetence. *Current Directions in Psychological Science, 12*(3), 83–87. https://doi.org/10.1111/1467-8721.01235

Eccles, J. (2009). Who am I and what am I going to do with my life? Personal and collective identities as motivators of action. *Educational Psychologist, 44*(2), 78–89. https://doi.org/10.1080/00461520902832368

Eccles, J. S., & Wigfield, A. (2020). From expectancy-value theory to situated expectancy-value theory: A developmental, social cognitive, and sociocultural perspective on motivation. *Contemporary Educational Psychology, 61*, 101859. https://doi.org/10.1016/j.cedpsych.2020.101859

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

Efklides, A. (2008). Metacognition: Defining its facets and levels of functioning in relation to self-regulation and co-regulation. *European Psychologist, 13*(4), 277–287. https://doi.org/10.1027/1016-9040.13.4.277

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
Ehrlinger, J., & Dunning, D. A. (2003). How chronic self-views influence (and mislead) estimates of performance. *Journal of Personality and Social Psychology, 84*, 5–17. https://doi.org/10.1037/0022-3514.84.1.5
Ehrlinger, J., Johnson, K., Banner, M., Dunning, D., & Kruger, J. (2008). Why the unskilled are unaware: Further explorations of (absent) self-insight among the incompetent. *Organisational Behavior and Human Decision Processes, 105*(1), 98–121. https://doi.org/10.1016/j.obhdp.2007.05.002
Epstein, R. M., Siegel, D. J., & Silberman, J. (2008). Self-monitoring in clinical practice: A challenge for medical educators. *Journal of Continuing Education in the Health Professions, 28*, 5–13. https://doi.org/10.1002/chp.149
Erickson, S., & Heit, E. (2015). Metacognition and confidence: Comparing math to other academic subjects. *Frontiers in Psychology, 6*, 742. https://doi.org/10.3389/fpsyg.2015.0074
Ernst, H. M., Wittwer, J., & Voss, T. (2023). Do they know what they know? Accuracy in teacher candidates' self-assessments and its influencing factors. *British Educational Research Journal, 49*(4), 649–673. https://doi.org/10.1002/berj.3860
Geraci, L., Kurpad, N., Tirso, R., Gray, K. N., & Wang, Y. (2023). Metacognitive errors in the classroom: The role of variability of past performance on exam prediction accuracy. *Metacognition and Learning, 18*, 219–236. https://doi.org/10.1007/s11409-022-09326-7
Gignac, G. E., & Zajenkowski, M. (2020). The Dunning-Kruger effect is (mostly) a statistical artefact: Valid approaches to testing the hypothesis with individual differences data. *Intelligence, 80*, 101449. https://doi.org/10.1016/j.intell.2020.101449
Gilovich, T., Griffin, D., & Kahneman, D. (2002). *Heuristics and biases: The psychology of intuitive judgements*. Cambridge University Press.
Guillory, J. J., & Blankson, A. N. (2017). Using recently acquired knowledge to self-assess understanding in the classroom. *Scholarship of Teaching and Learning in Psychology, 3*(2), 77. https://doi.org/10.1037/stl0000079
Holling, H., & Preckel, F. (2005). Self-estimates of intelligence—Methodological approaches and gender differences. *Personality and Individual Differences, 38*(3), 503–517. https://doi.org/10.1016/j.paid.2004.05.003
Hosein, A., & Harle, J. (2018). The relationship between students' prior mathematical attainment, knowledge and confidence on their self-assessment accuracy. *Studies in Educational Evaluation, 56*, 32–41. https://doi.org/10.1016/j.stueduc.2017.10.008
Kleitman, S., & Stankov, L. (2007). Self-confidence and metacognitive processes. *Learning and Individual Differences, 17*(2), 161–173. https://doi.org/10.1016/j.lindif.2007.03.004
Kleitman, S., Hui, J. S. W., & Jiang, Y. (2019). Confidence to spare: Individual differences in cognitive and metacognitive arrogance and competence. *Metacognition Learning, 14*, 479–508. https://doi.org/10.1007/s11409-019-09210-x
Komarraju, M., & Nadler, D. (2013). Self-efficacy and academic achievement: Why do implicit beliefs, goals, and effort regulation matter? *Learning and Individual Differences, 25*, 67–72. https://doi.org/10.1016/j.lindif.2013.01.005
Kruger, J., & Dunning, D. (1999). Unskilled and unaware of it: How difficulties in recognising 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
Magnus, J. R., & Peresetsky, A. A. (2022). A statistical explanation of the Dunning-Kruger effect. *Frontiers in Psychology, 13*, 840180. https://doi.org/10.3389/fpsyg.2022.840180
Marks, M. B., Haug, J. C., & Hu, H. (2018). Investigating undergraduate business internships: Do supervisor and self-evaluations differ? *Journal of Education for Business, 93*(2), 33–45. https://doi.org/10.1080/08832323.2017.1414025
McIntosh, R. D., Fowler, E. A., Lyu, T., & Della Sala, S. (2019). Wise up: Clarifying the role of metacognition in the Dunning-Kruger effect. *Journal of Experimental Psychology: General, 148*(11), 1882–1897. https://doi.org/10.1037/xge0000579
Miller, R. S., & Wang, M. (2019). Cultivating adolescents' academic identity: Ascertaining the mediating effects of motivational beliefs between classroom practices and mathematics identity. *Journal of Youth and Adolescence, 48*, 2038–2050. https://doi.org/10.1007/s10964-019-01115-x
Morony, S., Kleitman, S., Lee, Y. P., & Stankov, L. (2013). Predicting achievement: Confidence vs self-efficacy, anxiety, and self-concept in Confucian and European countries. *International Journal of Educational Research, 58*, 79–96. https://doi.org/10.1016/j.ijer.2012.11.002
Nagel, M., & Lindsey, B. (2018). The use of classroom clickers to support improved self-assessment in introductory chemistry. *Journal of College Science Teaching, 47*(5), 72–79.
OECD. (2015). *The ABC of gender equality in education: Aptitude, behaviour, confidence*. PISA, OECD Publishing.

Panadero, E., Brown, G. T., & Strijbos, J. W. (2016). The future of student self-assessment: A review of known unknowns and potential directions. *Educational Psychology Review, 28*, 803–830. https://doi.org/10.1007/s10648-015-9350-2

Parker, A. M., Bruine de Bruin, W., Yoong, J., & Willis, R. (2012). Inappropriate confidence and retirement planning: Four studies with a national sample. *Journal of Behavioral Decision Making, 25*(4), 382–389.

Peixoto, F., Radišić, J., Krstić, K., Hansen, K. Y., Laine, A., Baucal, A., Sõrmus, M., & Mata, L. (2023). Contribution to the validation of the Expectancy-Value Scale for primary school students. *Journal of Psychoeducational Assessment, 41*(3), 343–350. https://doi.org/10.1177/07342829221144868

Pekrun, R. (2006). The control-value theory of achievement emotions: Assumptions, corollaries, and implications for educational research and practice. *Educational Psychology Review, 18*, 315–341. https://doi.org/10.1007/s10648-006-9029-9

Pitkänen, H. (2023). Reflektoiva oppilas suomalaisessa perusopetuksen opetussuunnitelmallisessa arviointidiskurssissa [A reflective student in Finnish basic education curriculum evaluation discourse]. *Kasvatus ja aika [Education and Time], 17*(1), 50–73. https://doi.org/10.33350/ka.119400

Sanchez, C., & Dunning, D. (2018). Overconfidence among beginners: Is a little learning a dangerous thing? *Journal of Personality and Social Psychology, 114*(1), 10–28. https://doi.org/10.1037/pspa0000102

Schlösser, T., Dunning, D., Johnson, K. L., & Kruger, J. (2013). How unaware are the unskilled? Empirical tests of the "signal extraction" counter explanation for the Dunning-Kruger effect in self-evaluation of performance. *Journal of Economic Psychology, 39*, 85–100. https://doi.org/10.1016/j.joep.2013.07.004

Serra, M. J., & DeMarree, K. G. (2016). Unskilled and unaware in the classroom: College students' desired grades predict their biased grade predictions. *Memory and Cognition, 44*, 1127–1137. https://doi.org/10.3758/s13421-016-0624-9

Sheldon, J. O., Ames, R. D., & Dunning, D. (2014). Emotionally unskilled, unaware, and uninterested in learning more: Reactions to feedback about deficits in emotional intelligence. *Journal of Applied Psychology, 99*, 125–137. https://doi.org/10.1037/a0034138

Stankov, L. (2019). Applied metacognition and separation of confidence and accuracy in correlational studies. *Metacognition and Learning, 14*, 509–516. https://doi.org/10.1007/s11409-019-09212-9

Stankov, L., & Lee, J. (2008). Confidence and cognitive test performance. *Journal of Educational Psychology, 100*(4), 961–976. https://doi.org/10.1037/a0012546

Stankov, L., & Lee, J. (2014). Overconfidence across world regions. *Journal of Cross-Cultural Psychology, 45*(5), 821–837. https://doi.org/10.1177/0022022114527345

Stankov, L., & Lee, J. (2017). Self-beliefs: Strong correlates of mathematics achievement and intelligence. *Intelligence, 61*, 11–16. https://doi.org/10.1016/j.intell.2016.12.001

Stankov, L., Lee, J., Luo, W., & Hogan, D. J. (2012). Confidence: A better predictor of academic achievement than self-efficacy, self-concept and anxiety? *Learning and Individual Differences, 22*(6), 747–758. https://doi.org/10.1016/j.lindif.2012.05.013

Stankov, L., Morony, S., & Lee, Y. P. (2014). Confidence: The best non-cognitive predictor of academic achievement? *Educational Psychology, 34*(1), 9–28. https://doi.org/10.1080/01443410.2013.814194

Stankov, L. (1999). Mining on the "no man's land" between intelligence and personality. In P. L. Ackerman, P. C. Kyllonen, & R. D. Roberts (Eds.), *Learning and individual differences: Process, trait, and content determinants* (pp. 315–337). American Psychological Association. https://doi.org/10.1037/10315-014

Sullivan, P. J., Ragogna, M., & Dithurbide, L. (2019). An investigation into the Dunning-Kruger effect in sport coaching. *International Journal of Sport and Exercise Psychology, 17*(6), 591–599. https://doi.org/10.1080/1612197X.2018.1444079

Tashiro, J., Parga, D., Pollard, J., & Talanquer, V. (2021). Characterising change in students' self-assessments of understanding when engaged in instructional activities. *Chemistry Education Research and Practice, 22*(3), 662–682. https://doi.org/10.1039/D0RP00255K

Thawabieh, A. M. (2017). A comparison between students' self-assessment and teachers' assessment. *Journal of Curriculum and Teaching, 6*(1), 14–20. https://doi.org/10.5430/jct.v6n1p14

Vincent-Ruz, P. & Schunn, C. D. (2018). The nature of science identity and its role as the driver of student choices. *International Journal of STEM Education, 5*(48). https://doi.org/10.1186/s40594-018-0140-5

Vonkova, H., Papajoanu, O., Stipek, J., & Kralova, K. (2021). Identifying the accuracy of and exaggeration in self-reports of ICT knowledge among different groups of students: The use of the overclaiming technique. *Computers and Education, 164*, 104112. https://doi.org/10.1016/j.compedu.2020.104112

Walkowiak, S., Coutrot, A., Hegarty, M., Velasco, P. F., Wiener, J. M., Dalton, R. C., Hölscher, C., Hornberger, M., Spiers, H. J., & Manley, E. (2023). Cultural determinants of the gap between self-estimated navigation ability and wayfinding performance: Evidence from 46 countries. *Nature Scientific Reports, 13*, 10844. https://doi.org/10.1038/s41598-023-30937-w

Williams, T., & Williams, K. (2010). Self-efficacy and performance in mathematics: Reciprocal determinism in 33 nations. *Journal of Educational Psychology, 102*(2), 453–466. https://doi.org/10.1037/a0017271

Wollenschläger, M., Hattie, J., Machts, N., Möller, J., & Harms, U. (2016). What makes rubrics effective in teacher-feedback? Transparency of learning goals is not enough. *Contemporary Educational Psychology, 44*, 1–11. https://doi.org/10.1016/j.cedpsych.2015.11.003

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

Kajsa Yang Hansen
Department of Education and Special Education, University of Gothenburg, Sweden

Current themes of research:

Educational equity, teacher competence, sociology of education, socioeconomic status, educational measurement.

Relevant publications:

Yang Hansen, K., Radišić, J., Yi, D., & Liu, X. (2022). Contextual effects on students' achievement and academic self-concept in the Nordic and Chinese educational systems. *Large-scale Assess Educ,* 10, 16. https://doi.org/10.1186/s40536-022-00133-9.

Ding, Y., Yang Hansen, K., & Klapp, A. (2022). Testing measurement invariance of mathematics self-concept and self-efficacy in PISA using MGCFA and the alignment method. *European Journal of Psychology of Education*, 1-24.

Yang Hansen, K., Radišić, J., Liu, X. & Glassow, L.N. (2020). Exploring diversity in the relationships between teacher quality and job satisfaction in the Nordic countries: Insights from TALIS 2013 and 2018 In Frønes, T.S., Pettersen, A., Radišić, J., & Buchholtz, N. (Eds.), *Equity, Equality and Diversity in the Nordic Model of Education*, Springer.

Cecilia Thorsen
Department of and Behavioural Studies, University West, Sweden

Current themes of research:

Prerequisites for and results of education - equity in education, i.e., why students' background affects their results in school and schools' compensatory abilities.

Relevant publications:

Johansson, S., Yang Hansen, K. & Thorsen, C. (2023). A modeling approach to identify academically resilient students: evidence from PIRLS 2016. *European Journal of Psychology of Education*. https://doi.org/10.1007/s10212-023-00711-7.

Thorsen, C., Yang Hansen, K., & Johansson, S. (2021). The mechanisms of interest and perseverance in predicting achievement among academically resilient and nonresilient students: Evidence from Swedish longitudinal data. British Journal of Educational Psychology, 91(4), e12431. https://doi.org/https://doi.org/10.1111/bjep.12431 .

Thorsen, C., Johansson, S., & Yang Hansen, K. (2021). Does Learning English Require Grit? Examining the Differential Effects on Grades in Core Subjects. Journal For The Psychology Of Language Learning, 3(2), 61-85. https://www.jpll.org/index.php/journal/article/view/73 .

Jelena Radišić
Department of Teacher Education and School Research,
Faculty of Educational Science, University of Oslo, Norway

Current themes of research:

Motivation for learning, academic emotions, teacher competence, teacher beliefs and instructional practices

Relevant publications:

Radišić, J., Nortvedt, G. A., & Runde, R. K. (2023). Relationships Between Mathematics Self-Beliefs, Exposure to ICT In School, and Achievement on PISA 2012 Paper- and Computer-Based Mathematics Assessments. In C. Martin, B. Miller, & D. Polly (Eds.), *Technology Integration and Transformation in STEM Classrooms* (pp. 223-246). IGI Global. https://doi.org/10.4018/978-1-6684-5920-1.

Yang Hansen, K., Radišić, J., Ding, Y., & X. Liu (2022). Contextual effects on students' achievement and academic self-concept in the Nordic and Chinese educational systems. *Large-scale Assess Educ,* 10, 16. https://doi.org/10.1186/s40536-022-00133-9.

Radišić, J., Selleri, P., Carugati, F., & Baucal, A. (2021). Are students in Italy really disinterested in science? A person-centered approach using the PISA 2015 data. *Science Education*. https://doi.org/10.1002/sce.21611.

Yang Hansen,K., Radišić, J., Liu, X. & Glassow, L.N. (2020). Exploring diversity in the relationships between teacher quality and job satisfaction in the Nordic countries: Insights from TALIS 2013 and 2018 In Frønes, T.S., Pettersen, A., Radišić, J., & Buchholtz, N. (Eds.), *Equity, Equality and Diversity in the Nordic Model of Education*, Springer.

Radišić, J. & Baucal, A. (2018). Teachers' reflection on PISA items and why they are so hard for students in Serbia. *European Journal of Psychology of Education*. *33*(3), 445-466.

Francisco Peixoto
ISPA – Instituto Universitário, Center for Research in Education (CIE-ISPA), Lisboa, Portugal

Current themes of research:

Academic motivation, academic emotions, parental involvement in education, self-concept and self-esteem, teachers' identity, resilience and well-being

Relevant Publications:

Castro Silva, J., Peixoto, F., Galhoz, A., & Gaitas, S. (2023). Job demands and resources as predictors of well-being in Portuguese teachers. *European Journal of Teacher Education*. Advance online publication. https://doi.org/10.1080/02619768.2023.2288552.

Peixoto, F., Radišić, J., Krstić, K., Hansen, K. Y., Laine, A., Baucal, A., Sõrmus, M., & Mata, L. (2023). Contribution to the validation of the Expectancy-Value Scale for primary school students. *Journal of Psychoeducational Assessmen, 41*(3), 343-350. https://doi.org/10.1177/07342829221144868.

Forsblom, L., Pekrun, R., Loderer, K., & Peixoto, F. (2022). Cognitive appraisals, achievement emotions, and students' math achievement: A longitudinal analysis. *Journal of Educational Psychology, 114*(2), 346–367. https://doi.org/10.1037/edu0000671 .

Forsblom, L., Peixoto, F., & Mata, L. (2021). Perceived classroom support: Longitudinal effects on students' achievement emotions. *Learning & Individual Differences, 85,* 101959. https://doi.org/10.1016/j.lindif.2020.101959 .

Goos, M., Pipa, J., & Peixoto, F. (2021). Effectiveness of grade retention: A systematic review and meta-analysis. *Educational Research Review*, *34*, 100401, https://doi.org/10.1016/j.edurev.2021.100401 .

Anu Laine
Faculty of Educational Sciences, University of Helsinki, Finland

Current themes of research:

Motivation and affect in mathematics education, teaching and learning problem solving, students' mathematical knowledge and its development

Most relevant publications (in the field of psychology of education):

Vessonen, T., Hellstrand, H., Aunio, P., & Laine, A. (2023). Individual Differences in Mathematical Problem-Solving Skills among 3- to 5-Year-Old Preschoolers. International Journal of Early Childhood (2023). https://doi.org/10.1007/s13158-023-00361-2.

Peixoto, F., Radišić, J., Krstić, K., Yang Hansen, K., Laine, A., Baucal, A., Sõrmus, M., & Mata, L. (2023). Contribution to the validation of the Expectancy-Value Scale for primary school students. *Journal of Psychoeducational Assessmen, 41*(3), 343-350. https://doi.org/10.1177/07342829221144868.

Räsänen, P., Aunio, P., Laine, A., Hakkarainen, A., Väisänen, E., Finell, J., Rajala, T., Laakso, M-J., & Korhonen, J. (2021). Effects of Gender on Basic Numerical and Arithmetic Skills: Pilot Data From Third to Ninth Grade for a Large-Scale Online Dyscalculia Screener. Frontiers in education: educational psychology, 6, [683672]. https://doi.org/10.3389/feduc.2021.683672 .

Haataja, E., Salonen, V., Laine, A., Toivanen, M. & Hannula, M.S. (2021). The teacherstudent eye contact in interpersonal interaction during collaborative problem solving: A multiple gaze-tracking research in secondary mathematics education. Educational Psychology Review 33(1), 51-67. https://doi.org/10.1007/s10648-020-09538-w.

Holm, M., Björn, P. M., Laine, A., Korhonen, J., & Hannula, M. S. (2020). Achievement emotions among adolescents receiving special education support in mathematics. Learning and Individual Differences, 79, [101851]. https://doi.org/10.1016/j.lindif.2020.101851.

Xin Liu
Department of Teacher Education and School Research,
Faculty of Educational Science, University of Oslo

Current themes of research:

Instructional quality and teacher effectiveness, Motivation, Teacher well-being and career development, International large-scale analysis, Educational measurement and assessment, Cross-cultural comparative education

Relevant publications:

Liu, X., Valcke, M., Yang Hansen, K., & De Neve, J. (2022). Does school-level instructional quality matter for school mathematics performance? Comparing teacher data across seven countries. *Sustainability, 14*(9), 5267. https://doi.org/10.3390/su14095267 .

Yang Hansen, K., Radišić, J., Yi, D., & Liu, X. (2022). Contextual effects on students' achievement and academic self-concept in the Nordic and Chinese educational systems. *Large-scale Assess Educ,* 10, 16. https://doi.org/10.1186/s40536-022-00133-9 .

Zhou, N., & Liu, X. (2022). The Study in the Profiles of Secondary Teachers' Instructional Quality and their Relationship with Self-efficacy and Job Satisfaction: Based on the TALIS 2018 Shanghai Data. *Shanghai Research on Education*, 2022(1), 55-59.

Yang Hansen,K., Radišić, J., Liu, X. & Glassow, L.N. (2020). Exploring diversity in the relationships between teacher quality and job satisfaction in the Nordic countries: Insights from TALIS 2013 and 2018 In Frønes, T.S., Pettersen, A., Radišić, J., & Buchholtz, N. (Eds.), *Equity, Equality and Diversity in the Nordic Model of Education*, Springer.

Authors and Affiliations

Kajsa Yang Hansen[1] · **Cecilia Thorsen**[2] · **Jelena Radišić**[3] · **Francisco Peixoto**[4] · **Anu Laine**[5] · **Xin Liu**[3]

✉ Kajsa Yang Hansen
kajsa.yang-hansen@ped.gu.se

Cecilia Thorsen
cecilia.thorsen@hv.se

Jelena Radišić
jelena.radisic@ils.uio.no

Francisco Peixoto
fpeixoto@ispa.pt

Anu Laine
anu.laine@helsinki.fi

Xin Liu
xin.liu@ils.uio.no

1 Department of Education and Special Education, University of Gothenburg, Gothenburg, Sweden

2 University West, Trollhättan, Sweden

3 Department of Teacher Education and School Research, Faculty of Educational Science, University of Oslo, Oslo, Norway

4 Center for Research in Education (CIE-ISPA), ISPA – Instituto Universitário, Lisbon, Portugal

5 Faculty of Educational Sciences, University of Helsinki, Helsinki, Finland