Education and Information Technologies (2025) 30:14821–14851
https://doi.org/10.1007/s10639-025-13392-0

Cognitive biases in user experience and spreadsheet programming

Domicián Máté[1,2] · **Judit T. Kiss**[1] · **Mária Csernoch**[3]

Received: 13 September 2024 / Accepted: 20 January 2025 / Published online: 1 February 2025

Abstract

The impact of cognitive biases, particularly biased self-assessment, on learning outcomes and decision-making in higher education is of great significance. This study delves into the confluence of cognitive biases and user experience in spreadsheet programming as a crucial IT skill across various academic disciplines. Through a quantitative analysis, we investigate whether structured learning in spreadsheet programming can counteract self-assessment biases among higher education students. Specifically, our focus is on scrutinizing the accuracy of self-assessment in Excel proficiency among professional STEM students at the University of Debrecen, Hungary, by comparing traditional written and digital assessments. Our findings reveal that while high-achieving students tend to exhibit more accurate self-assessments, many students have a pervasive tendency to overestimate their spreadsheet competencies. These results emphasize the necessity for educational strategies that acknowledge cognitive biases in self-assessment, with far-reaching implications for curriculum design and lean management in higher education, by integrating evidence-based approaches to enhance digital competencies. This study makes a valuable contribution to the broader dialogue on improving learning outcomes and user experience in spreadsheet programming. Additionally, the research provides valuable insights for educators and policymakers, advocating for pedagogical adjustments that can assist students in better evaluating their skills and knowledge, thereby promoting more precise self-assessment practices.

Keywords Cognitive Bias · User Experience (UX) · Higher education · Lean management · Learning Outcomes · Spreadsheet Programming

Extended author information available on the last page of the article

1 Introduction

Cognitive biases are increasingly recognized as a significant, global public mental and health issue. Biased self-assessment, a crucial aspect of cognitive biases, refers to people's tendency to evaluate their abilities, traits, and characteristics in a skewed or inaccurate manner (Karpen, 2018). Moreover, biased self-assessment can lead to overconfidence in one's judgments and abilities, often resulting in underestimating the complexity of tasks or overestimating personal competence (Eva et al., 2018). The self-enhancement tendency involves viewing oneself in an overly favorable light, which boosts self-esteem but distorts self-assessment (Alicke & Sedikides, 2009). Another manifestation of this bias is the Dunning-Kruger effect, where less skilled individuals overestimate their abilities, while highly skilled ones may underestimate their competence (Dunning, 2011; Kruger & Dunning, 1999). This cognitive bias can impact decision-making, learning, and personal growth, as individuals may fail to recognize their strengths and weaknesses accurately (Kim et al., 2016).

The persistence of self-knowledge biases can be attributed to the fact that the mechanisms behind them operate at a subconscious level. Individuals possess various cognitive tools to downplay threatening information and emphasize positive information, with the most common tools being self-serving reasoning and biased hypothesis testing (Robins & Beer, 2001). As the psychological mechanisms driving biased self-assessment remain below conscious awareness, the direct strategies to address bias are unlikely to be successful (Kakinohana & Pilati, 2023). A more effective approach may involve structuring students' learning experiences in a way that hinders the efficient operation of these unconscious biasing mechanisms in higher education.

The research is centered on the relationship between end-user behavior and productivity in digital problem-solving. Previous studies have demonstrated that cognitive biases, overestimation, overconfidence, and errors are common in spreadsheeting, which results in significant losses (T. Nagy et al., 2021; Sarkar et al., 2020). While there are numerous tools available to correct spreadsheet errors, only some are dedicated to teaching error prevention and building up Computer Science (CS)-based spreadsheet knowledge. We advocate for a CS- and programming-focused approach to spreadsheeting, in line with the principles of the lean production system (Niklas & Pär, 2015). In support of our proposal, we analyzed various educational materials, including curricula, course descriptions, textbooks, help tools, printed and online tutorials, tests and exams (Csernoch et al., 2024a, 2024b). It is observed that these sources primarily emphasize tool-centered (low-mathability (Baranyi & Gilanyi, 2013), evidence-led innovation (Wolfram, 2020, 2024), push-production system approaches (Liker, 2004; Ohno, 1988; Rother, 2003, 2009; Womack & Jones, 1997) approaches and provide minimal coverage of CS and mathematical principles, with a focus on the interface, the operation, and functions of spreadsheet programs (Wolfram, 2020).

Despite the scientific findings of Computer Education Research (CER) (Csapó et al., 2020, 2021; Csernoch et al., 2024a, 2024b), there appears to be an increasing emphasis on hardware and software tools in spreadsheeting education. This challenge raises further concerns regarding biases inherent in digital education, particularly the issues of cascade and group noise and group polarization. We argue that there should be a more quantitative analysis of spreadsheet programming in end-user experience to address how the emphasis on tool-centered approaches might affect the understanding of CS and mathematical principles (Sarkar et al., 2020). Consequently, there is a need to assess the following research question. What evidence supports that the self-assessment of higher education students can be significantly improved through courses in spreadsheet programming?

The importance of this study is to address the observation that a significant proportion of tertiary students tend to inaccurately evaluate their level of spreadsheet knowledge, digital competence and skill (Csernoch et al., 2015; T. Nagy et al., 2021). Therefore, the aim of this study is to assess the proficiency of higher education students to evaluate their spreadsheet knowledge. A comparative analysis of traditional written and digital Excel tests will be conducted to analyze the impact of the identified factors on students' self-evaluation performance. However, there still needs to be more consensus on whether self-assessment skills can be enhanced during higher education courses (Lu et al., 2023).

The study explores whether high-achieving students have a more accurate self-assessment of their proficiency in spreadsheeting. The study was conducted at the University of Debrecen (Hungary), with a sample comprising students majoring in fields, informatics (pre-service teachers of informatics and library and information science), engineering, and economics. A rigorous quantitative analysis was employed to facilitate a comparative assessment of the present findings with those of the dominant research by employing a variety of statistical techniques, including linear and logistic regression, as well as non-parametric statistical analysis. This research paper sheds light on the cognitive biases inherent in self-assessment practices based on the problem-solving experiences of higher-education students in the context of spreadsheet programming. In the next sections, we present the literature background, data design and methodologies, and results. In conclusion, our findings offer concise insights that contribute to the ongoing empirical and proficient discourse on the enhancement of higher education in the domain of digital competencies.

2 Literature background

2.1 Self-evaluation literature

The literature pertains to studies that have explored how individuals assess their knowledge and abilities in relation to their proficiency in a specific area or in solving a particular problem (Jansen et al., 2021). These studies also look at the connection between an individual's self-assessment and their actual abilities (Moore & Healy, 2008). The Dunning-Kruger (DK) effect suggests that individuals with lower competence tend to overestimate their abilities and knowledge (Kruger & Dunning, 1999). One explanation for this phenomenon is the differences in metacognitive abilities (McIntosh et al., 2019), where individuals with lower capabilities may not be able to evaluate their performance accurately (Dunning et al., 2003).

The DK effect has been studied in various fields, such as mathematics and statistics (Hosein & Harle, 2018; Magnus & Peresetsky, 2022), computing (Gibbs et al., 2017), chemistry (MacNeil et al., 2024), economics, business (Feld et al., 2017; Sawler, 2021), financial literacy (Balasubramnian & Sargent, 2020; Gignac, 2022), health (Scheiber et al., 2023; Surdilović et al., 2022), accidents (Surdilović et al., 2022), information literacy (Jin et al., 2020; Mahmood, 2016; Nierenberg & Dahl, 2023), social psychology (Anson, 2018), and sports coaching ability (Sullivan et al., 2019). Alongside research confirming the DK effect, several studies are exploring its limitations. Some researchers even argue that the effect may be nothing more than a statistical artifact, while others question the correctness of the explanation (Hartwig & Dunlosky, 2014). Furthermore, the self-evaluation of individuals can be influenced by their previous experiences, beliefs, and newfound information gained (von Suchodoletz & Hepach, 2021). Consequently, there is a need to consider the effect of limited knowledge of abilities on the validity of the Dunning-Kruger effect (Lebuda et al., 2024).

The need for a clear consensus emphasizes further investigation of the DK effect. DK impact studies can be particularly important in addressing challenges and developing appropriate vaccination policies regarding childhood or pandemic anti-vaccination (Motta et al., 2018), as less informed individuals may overestimate their knowledge, and this may influence their decisions. Examining the impact of DK can be just as important from the point of view of developing an appropriate education policy (Schunk & Zimmerman, 1997). In instances where students have an inaccurate assessment of their proficiency in a specific subject, this can significantly impact their choices regarding education and dedication to self-enhancement (Chen et al., 2015). Conversely, accurate self-assessment among STEM (Science, Technology, Engineering and Mathematics) students preparing for exams can lead to an increase in career awareness and a reduction in dropout rates (Han et al., 2021).

Yang et al. (2024) demonstrated the impact of the Dunning-Kruger effect on the math results of Grades 3 and 4- students. They also found differences across six European countries in their comparative study. In addition to socioeconomic background and gender, the DK effect may be influenced by country-specific factors. However, in a similar study on mathematical performance, Hosein and Harle (2018) concluded that students with medium performance are more likely to evaluate themselves less accurately than students with low or high performance.

The deviation of self-evaluation from objectively measured performance, such as under- or overvaluation, is changed with age. Xia et al. (2024) investigated whether children tend to overestimate their abilities in performing certain tasks. However, the study observed that children's metacognitive ability develops with age, leading to a decrease in overestimation from early to late childhood. In contrast, Prims and Moore (Prims & Moore, 2017) did not find evidence to support the idea that younger people tend to overestimate their abilities in their study of individuals aged 18–75. Crawford and Stankov (1996) revealed an ineffective but positive relationship between overconfidence and age for intelligence. Schaefer et al. (2023) explored the relationships between individuals' perceptions of task difficulty and their actual performance across different age groups and genders. Among the four groups separated by age (children, teenagers, younger adults, and older adults), young adults judged the difficulty level of the tasks most accurately in relation to their performance. Pak and Chatterjee (2016) showed that older people overestimate their abilities and skills related to financial investment, and the degree of overestimation increases with age. However, in a health-related Chinese study, Jia et al. (2023) determined that older adults overestimated their health status and underestimated the risks associated with smoking. Meanwhile, Gignac (2022) studied people older than 18 and found that there was no age difference in the Dunning-Kruger effect related to financial literacy. The author also could not confirm the impact of DK (Gignac, 2024).

Prompted by the question of whether the Dunning-Kruger effect can be observed and if it changes with age as a result of learning and experience, we are led to investigate the importance of time spent in higher education or the completion of individual courses from the perspective of the DK effect. De Bruin et al. (de Bruin et al., 2017) found that college students' accuracy in predicting their exam scores decreased as the course progressed and their ability improved. This phenomenon led to a reduction in the overestimation of lower-performing students, while the underestimation of higher-performing students was observed compared to the actual results. Gonda (2022) surveyed environmental awareness and conscious consumer behavior in relation to sustainable development. The sample was grouped into six categories based on their education levels: elementary school, vocational school, vocational high school, high school, tertiary technical school, college and university degree. The results revealed the DK effect, indicating that it decreases as the education level increases. In simpler terms, individuals with higher education levels tended to assess their knowledge more accurately.

Several empirical studies have examined the performance of high school and university students (Anson, 2018; de Bruin et al., 2017; Gonda, 2022; Nierenberg & Dahl, 2023). In higher education, researchers typically engage in survey research pertaining to the bachelor's and master's degree levels (Kun, 2016; Nepal & Kafle, 2024). Nierenberg and Dahl (2023) examined the accuracy of students' self-assessment of their skills before and after taking tests. The results revealed that students' information literacy skills improved as they advanced in education. Prior to the tests, first-year and master students tended to overestimate their abilities, while PhD generally underestimated their skills. After completing the tests, all groups of students underestimated their performance despite showing improvement. Surdilović et al. (2022) examined final-year dental students' communication, diagnostic, and clinical skills for an entire academic year. Based on the results, they found that in addition to the improvement in ability, the students' overestimation of their ability did not show a decrease since their self-confidence decreased. When comparing abilities related to a specific research field and self-assessment, the survey results can be misleading. For example, classifying students as better or less able based on exam results may not confirm the accuracy of their self-assessment, as preparation can also impact results.

2.2 Spreadsheet programming

In the realm of spreadsheet software (Abraham et al., 2009; Sestoft, Wakeling, Sarkar et al., Burnett et al., 2001), there is a prevalent belief that spreadsheets provide a platform for programming that closely adheres to the principles of functional programming (Booth, 1992), effectively serving as a simplified functional language. However, neither computer science nor education discipline appears to facilitate this advantage of spreadsheeting (Wolfram, 2020). The focus of both is primarily tool-centered, with an emphasis on the environment rather than problem-solving approaches. In contrast, our research group has developed Sprego (Spreadsheet Lego) (Csernoch, 2014), which is based on the programming paradigm hidden behind spreadsheets and Booth's findings regarding the role of functional programming as a first language (Booth, 1992).

Sprego meticulously aligns with the foundational principles of computer science, incorporating Cognitive Load Theory (Sweller et al., 2011) and cognitive processes (Kahneman, 2011). Moreover, it embraces the concept-based problem-solving methodology (Polya, 1945) and embodies the long-term thinking philosophy of the Toyota Production System (TPS) (Ohno, 1988). The strong integration of various theoretical frameworks highlights the resilience of the Sprego approach. The learning-teaching approach utilizes real-world data for comprehensive analysis, mainly sourced from the internet or private collections (Csapó et al., 2021; Csernoch, 2014). It facilitates the development of a kanban system, starting from the desired output requirements and formulating algorithms (plans) based on the needs. The coding is performed using a minimal set of spreadsheet functions (Csernoch, 2014). Emphasis is placed on scrutinizing the problem-solving process and the outputs of the algorithm steps (builds automation with a human touch, e.g., jidoka). If an error is detected during the process, it is halted and not resumed until the error is rectified. An important aspect of Sprego is its ability to differentiate between errors and error-outputs, while also instructing students on how to recognize genuine errors and correct results with error-outputs.

In accordance with Sprego's Just-in-Time approach (Csernoch et al., 2024a, 2024b), which emphasizes the use of tools tailored to the specific task at hand, it is apparent that the focus lies in utilizing only the necessary teaching instruments to address immediate issues. The teacher is primarily responsible for the selection, considering the background knowledge, skills, competencies, interests, age, and other school subjects of the students to provide an appropriate learning experience (Kim et al., 2019). The chosen problems should give students enough time and space to practice and develop mental frameworks for long-term memory storage. The concept of mental frameworks is crucial in education and effective problem-solving (Kahneman, 2011; Sweller et al., 2011), as they enable fast thinking when needed, enhancing their efficiency and effectiveness. In the absence of mental frameworks, a slower, more deliberate thought process becomes imperative, thereby elevating the likelihood of errors (Kahneman, 2011; Panko, 2013). Additionally, if misconceptions are stored, they can lead to incorrect results, potentially influenced by end-user confidence (Abraham et al., 2009; Panko, 2013; Sarkar et al., 2020; Wakeling, 2007).

Table 1 The fundamental set of sprego functions

Sprego text	Sprego numbers	Sprego pro
LEFT()	SUM()	IF()
RIGHT()	MIN()	MATCH()
LEN()	MAX()	INDEX()
SEARCH()	AVERAGE()	ISERROR()

Source: (Csernoch, 2014; Csernoch & Bíró, 2015)

Sprego stands out due to its emphasis on the concept of simplified functional programming (Table 1). This approach draws from mathematical concepts while providing practical applications. The spreadsheet offers students a valuable opportunity to grasp the practical benefits of functions, extending beyond their traditional use for graphing in math class (Burnett et al., 2001). Sprego not only incorporates the set of the limited number of functions but also expands upon it, enabling the manipulation of various data types, n-ary, and multilevel functions. These advanced concepts surpass the boundaries of elementary mathematics.

As noted by Wolfram, in a modern computer-based (math) education environment, the curriculum can be organized based on conceptual complexity rather than computational complexity (Wolfram, 2020). Furthermore, Sprego enables the construction of multilevel functions. In this formula, the functions are embedded within one another, e.g., INDEX(CAPITAL,MATCH(MAX(AREA),AREA,0)). This feature enhances students' comprehension of how the output of one function can serve as the input for another. In essence, Sprego offers a harmonious blend of constructionism and targeted instruction (Csapó et al., 2020, 2021; Csernoch, 2014).

2.3 Cognitive biases influencing spreadsheet programming

Cognitive biases are systematic patterns of deviation from norm or rationality in judgment, which can significantly impact user experience (UX) and spreadsheet programming (Blanco, 2017). While the Dunning-Kruger effect – where individuals with low ability overestimate their competence – is notable (Kruger & Dunning, 1999), a comprehensive understanding of various cognitive biases is essential for creating effective teaching environment.

The functions offered by Sprego (Table 1) touch upon a notable cognitive bias known as the less-is-more pattern (Kahneman, 2011). This collection of Sprego functions is developed from algorithmic and programming principles, utilizing insights from cognitive load research (Sweller et al., 2007, 2011). One pressing question for Sprego and similar programming and computer science environments (Hubwieser, 2004) is how to integrate new functionalities into an existing set. The less-is-more patterns suggest that while larger sets are often viewed more favorably during joint evaluations, they tend to be less appreciated in individual evaluations (Kahneman, 2011). This phenomenon can be attributed to the behavior of fast thinking (System 1), which typically averages responses during single assessments. It implies that educators, learning materials, and end-user interfaces may continuously add and propose new features, provided there is an opportunity for slow thinking (System 2) to engage. However, it has been observed that System 2 is prone to errors, underscoring the importance of allowing System 1 to operate to reduce mistakes (Panko, 2013). Panko (2013) observed that slow thinking is often applied during spreadsheet activities, which can lead to errors. Based on these findings, reducing cognitive load through a high-mathability approach allows both System 1 and System 2 to contribute effectively to problem-solving in spreadsheeting.

Kahneman (2011) notes that if they added a cheap gift to the expensive set of products that made the whole deal less attractive. Less is more in this case. These findings lend credibility to Sprego. The set is inherently limited due to cognitive load, but it can be expanded in alignment with fundamental concepts. This result also sheds light on our earlier finding that tool-centered and algorithm-driven spreadsheeting cannot be combined effectively (Csernoch et al., 2024a, 2024b). Furthermore, it implies that continuously expanding the functionalities of digital products may not be sustainable as the cognitive load on end-users increases, making it challenging for them to fulfill their designated roles (Csernoch et al., 2024a; Nagy & Csernoch, 2023). This outcome aligns with Collins' findings, which highlight that organizations seeking to progress from good-to-great must adhere to a clear concept (Collins, 2001). If an opportunity does not align with this framework, it should be rejected. For example, if new features or functions for Sprego do not resonate with our core concept, we must forget them.

Additionally, the noise effect refers to the psychological phenomenon where individuals adopt beliefs or behaviors based on the actions of others (Kahneman et al., 2021), often without critical evaluation (Blanco, 2017). In the realm of spreadsheet programming, this can manifest as the inclination to embrace widely utilized functions or formulas simply because they are popular, rather than assessing their applicability or efficiency for a given context. This tendency may lead to suboptimal decision-making or the perpetuation of inefficiencies in data analysis and modeling. This confirmation bias is the tendency to seek, interpret, and retain information that reinforces one's preconceptions. It can lead users to prioritize feedback that supports their initial ideas, potentially overlooking critical user needs (Frank et al., 2024). The social desirability bias refers to the tendency of respondents to answer questions in a way that will be viewed favorably by others (Azzopardi, 2021; Kahneman et al., 2021). In the context of this research, it may lead participants to provide feedback that they believe the researcher expects rather than expressing their authentic opinions, thereby skewing the collected data. Similarly, the false consensus effect is the inclination to overestimate the extent to which others share our beliefs and behaviors (Friedman, 2023). Recognizing and addressing these cognitive biases influencing decision-making is essential for both professionals and spreadsheet programmers. By understanding and mitigating the impact of these biases, practitioners can make more informed decisions, ultimately enhancing user experiences and improving data management accuracy.

2.4 Cascade and group noise and group polarization in digital education

End-user computing remains one of the most controversial topics within the field of Computer Science (CS). CS is notably shaped by various stakeholders, e.g., developers, traditional computer scientists, industry professionals, educational institutions, and the end-users themselves. Given this landscape, it is unsurprising that cognitive biases emerge from these influences. Each of these participants seeks to advance their interests, whether intentionally or intuitively. Kahneman et al. (2021) have established that both cascade and group noise play a critical role in decision-making. As noted by Kahneman et al., there are 'wise crowds', whose mean judgment is close to the correct answer, but there are also crowds that follow tyrants, fuel market bubbles, believe in magic, or are under the sway of a shared illusion.

The low-efficiency of end-users can also be explained by the dynamics among group members where the level of noise is high (Carroll, 1987; Kirschner & De Bruyckere, 2017; Rattenbury et al., 2017; Reynolds, 2008). A great deal depends on social influences and on whether people see that other people are attracted or repelled. Kahneman et al. (2021) suggest that the perceived consensus among group members can be misleading, as it may stem from the initial opinions of a limited number of individuals rather than genuine collective wisdom. Additionally, they highlight those informational cascades can steer groups toward significantly detrimental outcomes.

Digital education is among the fields that are significantly affected by group and cascade noise. A notable contributor to this noise is termed'folk pedagogy', like folk medicine, negatively impacts digital education (Lister, 2008). Malmi et al. (2019) asserted that practice-based papers predominate in the discipline. However, a substantial amount of work is being conducted in Computer Education Research (CER), and the current influence of theoretical constructs on ongoing research appears to be limited (Malmi et al., 2023). If folk pedagogy overshadows CER, it can generate significant noise, leading research in misguided directions.

Moreover, the Dunning-Kruger effect contributes to the problem when under-qualified participants voice their opinions first, and it can evoke cascade noise. Given that end-user computing operates within a vast industry of varied interests, participants are eager to share their views, further amplifying the noise. This noise can result in group noise and group polarization, where individuals typically shift toward more extreme positions in line with their initial inclinations after interacting with one another.

Additionally, if individuals are concerned about their reputation within the group, they may gravitate toward the dominant perspective, further contributing to polarization (Kahneman et al., 2021). If people find their convenience more appealing (either conscious or not) compared to group interest, it can lead to the same result. Group polarization can lead to errors if the inefficiency of end-user activities is accepted and the errors within digital artifacts are tolerated and endorsed by the educational systems. Then, teachers are likely to adopt this mindset, which, in turn, influences students. The noise generated by folk pedagogy, the Dunning-Kruger effect, and the perspectives of billions of end-users, along with outsiders, contribute to group polarization and render end-user computing one of the less sustainable areas within Computer Science and its applications. This bias needs to be acknowledged, and transformative changes are necessary to reduce the noise surrounding digital education.

3 Materials and methods

This research utilized a quasi-experimental design to assess the impact of different teaching methodologies on the performance of professional end-users—specifically, STEM-oriented university students at the University of Debrecen, Hungary. The study compared groups that engaged with the Sprego method, traditional tool-centered approaches, and a combination of both.

3.1 Research design

The study involved 173 students enrolled in bachelor's (BSc) and master's (MSc) programs across the disciplines of Economics and Business, Engineering, and Informatics. The sample comprised 69 master students and 104 bachelor students. In terms of language of instruction, there were 66 international students and 107 Hungarian students.

Participants were categorized into three groups based on the instructional method they received:

- Sprego Group (DPULL): This group consisted of 27 students who were taught using the Sprego methodology. Sprego is a deep-approach programming tool designed for spreadsheet environments, focusing on general-purpose functions to enhance algorithmic, programming and problem-solving skills.
- Non-Sprego Group (DPUSH): This group included 107 students who underwent traditional, tool-centered instructional methods.
- Mixed-Study Group (DMIXED): Comprised of 39 students, this group experienced a blend of both teaching methodologies.

In this quasi-experimental research framework, the independent variables were the teaching methodology (Sprego, traditional, or mixed), programs and internalization. In contrast, the dependent variables were the students' performance and assessment outcomes. Unlike true experimental designs, quasi-experiments do not utilize random assignment, which can lead to selection biases. These factors may impact the internal validity of the study. Nevertheless, they offer valuable insights into educational contexts where randomization may not be feasible (Handley et al., 2018).

Data were gathered through assessments that measured students' proficiency and understanding of the material covered. The analysis aimed to evaluate the effectiveness of the Sprego methodology compared to traditional teaching methods. Previous studies have suggested that Sprego can be more effective than traditional tool-centered approaches in teaching spreadsheet programming (Csapó et al., 2020, 2021).

	A	B	C	D	E	F	G
1	Country	Continent	Capital	Area	Population (thousand)		
2	Afghanistan	Asia	Kabul	647500	27756		
3	Albania	Europe	Tirana	28748	3545		
4	Algeria	Africa	Algiers	2381740	32278		
5	American Samoa	Oceania	Pago Pago	199	69		
6	Andorra	Europe	Andorra la Vella	468	68		
7	Angola	Africa	Luanda	1246700	10593		
8	Anguilla	Amerika	The Valley	102	12		
233	Yemen	Asia	Sanaa	527970	18701		
234	Yugoslavia	Europe	Belgrade	102350	10657		
235	Zambia	Africa	Lusaka	752614	9959		
236	Zimbabwe	Africa	Harare	390580	11377		

Fig. 1 Sample table of paper-based testing Source: author's compilations

3.2 Testing process

During the testing process, two distinct test versions were employed: a paper- and an Excel-based version. Potential differences between the versions were accounted for in the analyses, with these variations considered separately.

The testing process comprised three major phases.

- Self-assessment

Evaluations of self-assessment can be divided based on whether individuals measure their competence in an absolute or relative manner. In absolute evaluation, individuals assess their abilities against an objective standard. In relative evaluation, they compare their knowledge to their previous understanding or the performance of others. In the initial phase, participants are required to estimate their absolute proficiency in spreadsheet knowledge on a scale of 0–100%.

- Paper-based testing

Subsequent to the self-evaluation phase, a spreadsheet test is presented in paper form, accompanied by a sample table (Fig. 1). The table comprises five fields, a row of field names, and 235 records. The data type of Fields Country, Continent, and Capital are string, whereas Area and Population (thousand) are integers. It should be noted that rows 9–232 are not visible in the presented figure. Additionally, the figure indicates that Cells G2 and G3 serve as variables, which can assume any value that is suitable for solving the presented problems.

The participants were presented with a test comprising seven spreadsheet tasks (Tasks 1–7). In Tasks 1–5, the participants were required to answer using a spreadsheet formula (Fig. 2).

The description of the tasks employs the use of the algorithms and data sources as follows. Task 1 is a linear search, which can be solved through the utilization of the MAX(), MATCH(), and INDEX() functions in conjunction with one another. The

1. What is the capital city of the largest country?

....

2. What is the population density of each country?

....

3. How many African countries are in the table?

....

4. What is the average population of those countries whose surface area is smaller than G2?

....

5. How many countries have a surface area greater than G3?

....

Fig. 2 Tasks 1–5 of paper-based testing Source: author's compilations

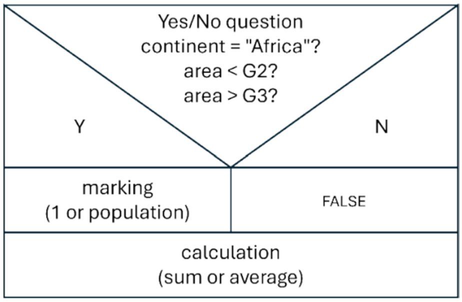

Fig. 3 The algorithm of conditional calculations (Tasks 3–5) Source: author's compilations

objective of Task 2 is to calculate the population density of each country. Participants must determine the numerator and denominator, taking care to note the field name of the population, which indicates that the data are given in thousands. Tasks 3–5 are conditional calculations, sharing the same algorithm but requiring different levels of abstraction. Task 3 presents the simplest level of abstraction, requiring the application of a conditional counting process with a constant value serving the condition. Task 4 represents the highest level of abstraction, wherein an inequality must be evaluated with a variable, and the average of the selected values must be calculated. Task 5 represents a level of abstraction between the other two, comprising a conditional counting process that requires the checking of an inequality with a variable.

The distinguishing feature of these three tasks (Tasks 3–5), which employ the same algorithm (Fig. 3), is that the syntax differs when built-in functions are used (COUNTIF(), COUNTIFS(), AVERAGEIF(), AVERAGEIFS(), SUMIF(), SUMIFS(), DCOUNT(), DAVERAGE(), DSUM()). It is due to the presence of equality versus inequality and constant versus variable. Conversely, if the algorithm of the conditional calculation is constructed and coded, there is no distinction between the solutions of the three tasks.

6. What is the result of the following formula?
{=SUM(IF(B2:B236="Europe",IF(LEFT(A2:A236)="A",1)))}

....

7. What is the result of the following formula?
{=SUM((B2:B236="Europe")*(LEFT(A2:A236)="A"))}

....

Fig. 4 Tasks 6–7 of paper-based testing Source: author's compilations

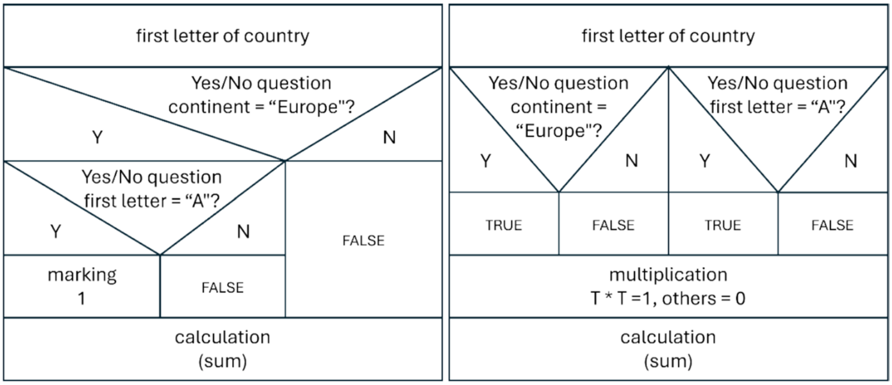

Fig. 5 The algorithms of conditional calculations (Tasks 6–7) Source: author's compilations

In contrast, in Tasks 6–7, the participants were asked to guess the output of the formulas and to provide answers in the form of natural language sentences. Moreover, an analysis of the formulas can reveal the algorithm, which could have facilitated the solution if interpreted correctly (Fig. 4).

Additionally, the two formulas yield the same output, which is the number of European countries where the country name begins with the letter A. The algorithm hinges on the premise that if the response to the yes/no question is negative, the output value is FALSE, which is excluded from the calculation of the sum (Fig. 5).

Furthermore, Tasks 6–7 and Tasks 3–5 share the same conditional calculation algorithm, which implies that understanding Tasks 6–7 can help to solve Tasks 3–5.

- Excel-based testing

In the third phase of the testing process, participants were required to complete the same Tasks 1–5 in an Excel environment, which suggests that they had already acquired familiarity with the tasks. The participants were provided with

	A	B	C	D	E	F	G	H
1	Country	Continent	Capital	Area	Population (thousand)	population density		
2	Afghanistan	Asia	Kabul	647500	27756			
3	Albania	Europe	Tirana	28748	3545			
4	Algeria	Africa	Algiers	2381740	32278			
5	American Samoa	Oceania	Pago Pago	199	69			
6	Andorra	Europe	Andorra la Vella	468	68			
7	Angola	Africa	Luanda	1246700	10593			
8	Anguilla	Amerika	The Valley	102	12			
233	Yemen	Asia	Sanaa	527970	18701			
234	Yugoslavia	Europe	Belgrade	102350	10657			
235	Zambia	Africa	Lusaka	752614	9959			
236	Zimbabwe	Africa	Harare	390580	11377			

Fig. 6 Sample table of Excel-based testing

a PDF document, accessible throughout the process, which listed the tasks and included a modified figure to assist them in entering the formulas into the appropriate cells (Fig. 6). Cells G2 and G3 were designated as variables, as in the paper solution (Fig. 7). However, in Excel, the participants were permitted to enter any data they deemed appropriate for evaluating the accuracy of their formulas.

3.3 Ethical issues

The testing process was carried out during the semester in classes that require spreadsheet knowledge and is aligned with the students' curricula. The students were informed of the nature, purposes, and process of the testing and consented to take part in it. The test comprised three distinct phases, necessitating the identification of participants. For identification purposes, the students' university IDs were used solely to align different sources. Aside from this use of IDs, the testing is anonymous; the students cannot be identified in the analysis or publication phase.

3.4 Evaluation of the tasks

The primary objective of the evaluation is to categorize all responses into the most granular units, each of which is assigned a value of one. The total number of points awarded for a given task is converted to a percentage. The resulting percentages were aggregated to calculate the overall test score, which was then compared to the self-evaluation value collected during the initial phase of the test.

The tasks presented have multiple correct solutions. These solutions were broken down into items, evaluated, and converted to percentages to ensure comparability of results. Both single and multilevel formulas were accepted if their sources were properly indicated. Moreover, named ranges were also accepted without any indication of their intentional use, both on paper and in Excel. If an algorithm/formula was expressed in natural language on paper, it was deemed acceptable and interpretable. In Excel, if cells G2 and G3 were blank and the

1. What is the capital city of the largest country?
   (Cell H1)

   ………………………………………………………………………………………

2. What is the population density of each country?
   (Column F)

   ………………………………………………………………………………………

3. How many African countries are in the table?
   (Cell H3)

   ………………………………………………………………………………………

4. What is the average population of those countries whose surface area is smaller than G2?
   (Cell H4)

   ………………………………………………………………………………………

5. How many countries have a surface area greater than G3?
   (Cell H5)

   ………………………………………………………………………………………

Fig. 7 Tasks 1–5 of Excel-based testing

formula produced an error but was correct or partially correct (recognizable items were present), the result was deemed valid and assessed.

3.5 Hypotheses statement

Based on previously published results influenced by the DK effect (Dunning, 2011; Kruger & Dunning, 1999; Máté & Darabos, 2017), we hypothesized that higher-achieving students are more accurate in assessing their spreadsheet knowledge than lower-achieving students. The extent of discrepancies between self-assessments and test results will serve as the basis for evaluating the digital competency instruction. We introduce the following hypotheses. In the first phase of the evaluation, students' self-assessment values and their achievement on the test are compared. Based on previously published results, the following hypothesis is formed:

> **H1:** Higher-achieving students predict their spreadsheet knowledge more accurately, as measured by the absolute value of the pre-examination assessment results, than their lower-achieving ones.

Former studies also consider whether more educated master (MSc) students estimate their performance more precisely than bachelor (BSc) ones (Dunning et al., 2004; Kruger & Dunning, 1999). The following hypothesis is formed:

Table 2 Statistical characteristics of the pre-estimated and actual (Paper and Excel) test scores

Scores	Min	Max	Mean	STD	Skewness	Kurtosis	Shapiro–Wilk Test
Student	2	100	47.63	21.66	−0.083	−0.688	0.977***
Paper	0	95	35.44	22.41	0.244	−0.621	0.975***
Excel	0	100	35.12	24.41	0.531	−0.262	0.960***

Source: author's estimations

N = 173, Significance p-levels: ***: 1 percent, **: 5 percent, *: 10 percent.

H2: In comparison to bachelor (BSc) students, master (MSc) students are more proficient at accurately predicting their spreadsheet knowledge.

Furthermore, it is assumed that students who have learned spreadsheets with the Sprego programming approach predict their examination results more accurately than students without spreadsheet programming experience. Students who are not proven Sprego learners form the control group. It is assumed that the control group was primarily taught using widely accepted tool-centered learning-teaching approaches (Gibbs et al., 2017).

H3: Students with experience in Sprego programming predict their spreadsheet knowledge more accurately than students without Sprego experience.

3.6 Multivariate regression models

In order to show that high-achieving students evaluate their exam results more accurately than lower-achieving students, we need to conduct advanced multivariate regression models. The upcoming linear models (Eq. 1) will include dependent variables that reflect the accuracy of students' estimations before the exam.

$$\begin{aligned} PREDIF_i = {} & \beta_o + \beta_1 FINSC_i + \beta_2 DMAJOR_i + \beta_3 DPULL_i \\ & + \beta_4 DMIXED_i + \beta_5 DPUSH_i + \beta_6 DLANGUAGE_i + \varepsilon_i \end{aligned} \tag{1}$$

When evaluating hypotheses H1, H2, and H3, we employ linear regression models with the dependent variable representing the accuracy of students' pre-estimations (PREDIF), determined by the absolute difference between student-assessed spreadsheet knowledge and the tutor-assigned test scores.

FINSC is replaced by tutor-assigned test scores of distinct exam types as independent variables. In the regression models, DMAJOR (MSc = 1, BSc = 0) dummy variables control the students' graduation levels. Students were assigned values if they studied based on Sprego (DPULL = 1 otherwise 0), if they studied with traditional, tool-centered methods (DPUSH = 1 otherwise 0) and if they studied mixed (DMIXED = 1 otherwise 0). ε denotes the error term at time i. The designation of DLANGUAGE (International = 1, Hungarian = 0) is intended to facilitate the internalization of higher education, thereby optimizing the goodness of fit, assessed by

R-squared (R^2) as the variation of the estimated variable around the sample mean compared to the variation of the observed variable.

Table 2 presents comprehensive descriptive statistics (minimum, maximum, mean, standard deviation (STD), and distribution measures) of test variables, including students' and tutors' assessments. The results of the skewness and kurtosis calculations indicated that the sample was not symmetric and did not exhibit a peak. Therefore, the sample values cannot be normally distributed. The Shapiro–Wilk test indicates that the observed significant values are unlikely to have been sampled from a normal distribution. Consequently, the test scores are not normally distributed in the sample population. The results will subsequently be employed to elucidate the elements under consideration in the models.

An additional regression analysis, which directly focuses on self-estimation of the degree of cognitive biases, implies that lower-performing students tend to overestimate their knowledge. In order to examine the correlation between student performance and the precision of their self-assessment of spreadsheet proficiency, a binary logistic regression model can be an appropriate analytical tool, provided that other variables remain constant (see Eq. 2). OPREDIF refers to a dependent (binary) variable denoted as Overestimate = 1 and 0 otherwise. It represents the discrepancy between a student's self-assessment values and scores in their test results. In the context of an overestimation, the student's confidence in their spreadsheet knowledge is higher than the actual outcome. The independent variables remain unchanged from the previous iteration.

$$P(OPREDIF)_i = \frac{e^{\beta_o+\beta_1 FINSC_i+\beta_2 DLANGUAGE_i+\beta_3 DMAJOR_i+\beta_4 DPULL_i+\beta_5 DMIXED_i+\beta_6 DPUSH_i}}{1+e^{\beta_o+\beta_1 FINSC_i+\beta_2 DLANGUAGE_i+\beta_3 DMAJOR_i+\beta_4 DPULL_i+\beta_5 DMIXED_i+\beta_6 DPUSH_i}} \tag{2}$$

4 Results of the estimations

4.1 Testing the hypotheses

The result of the statistical analysis presented in Table 3 demonstrates significant relationships between students' prediction accuracy and test results in both models. Consequently, the accuracy of self-assessment is independently estimated in two models. The initial model (PAPER) encompasses all available independent variables, whereas the alternative model (EXCEL) is confined to those significant at p-levels.

The impact of test results on the discrepancy between self-assessment and test results appears to be relatively modest, with coefficients between − 0.843 and − 0.787 (See Table 3, Model 1–4). Accordingly, there is a negative correlation between students' outcomes and the accuracy of their assessments. Hence, the results confirm the H1 hypothesis that higher-achieving students are more adept at predicting their examination results than their lower-achieving peers.

Table 3 Results obtained from the linear regression models for the preliminary self-assessment examination

Dependent variable	PREDIF			
Type of tests	PAPER		EXCEL	
Independent Variables/Models	Model 1	Model 2	Model 3	Model 4
Constant	39.784 (9.893)***	32.408 (8.310)***	40.749 (10.632)***	33.998 (9.868)***
FINSC	−0.787 (−10.580)***	−0.823 (−11.419)***	−0.810 (10.632)***	−0.843 (−13.091)***
DMAJOR	7.504 (1.994)**	6.960 (1.844)*	6.935 (1.832)*	6.491 (1.708)*
DPULL	−14.297 (−2.824)***		−13.490 (−2.705)***	
DMIXED	−5.140 (−1.128)		−4.210 (−0.923)	
DPUSH		8.913 (2.211)**		8.175 (2.049)**
DLANGUAGE	1.961 (0.468)	1.671 (0.396)	0.957 (0.231)	0.826 (0.198)
R^2	0.532	0.523	0.547	0.539
Adjusted R^2	0.518	0.512	0.533	0.528
Durbin-Watson	1.998	1.939	2.023	1.954
Max(VIF)	1.551	1.622	1.551	1.587

Source: author's estimations

The t-statistics reported in parentheses are robust to heteroscedasticity. Letters in the superscript indicate significance levels: ***: 1 percent, **: 5 percent, *: 10 percent. P-values without a superscript suggest that the coefficient is insignificant even at the 10 percent level.

Ceteris paribus, master students are less likely to predict their spreadsheet knowledge accurately than their undergraduate colleagues (the H2 hypothesis was rejected). More interestingly, we found that students with experience in Sprego programming (DPULL) seem to predict their spreadsheet knowledge more accurately than students without Sprego experience (DPUSH). The H3 hypothesis can be accepted. Furthermore, there are no observable language differences.

Variance Inflation Factor (VIF) is a statistical measurement that quantifies the severity of multicollinearity in a regression analysis and measures how much the variance of the estimated regression coefficients is inflated in the model. Generally, the VIF values above 5 indicate a problematic multicollinearity level among the regression model's independent variables (Hair, 2010). The Durbin-Watson (D-W) measure is widely regarded as the most powerful method for assessing first-order autocorrelation. D-W test values within the 1.5–2.5 range are deemed normal (Turner, 2020).

Table 4 Results obtained from the binary logistic regression models for the self-assessment examination

Dependent variable	OPREDIF							
Type of tests	PAPER				EXCEL			
Independent Variables	Model 1		Model 2		Model 3		Model 4	
	β	Exp(β)	β	Exp(β)	β	Exp(β)	β	Exp(β)
Constant	5.162 (34.92)***	174.575	3.364 (27.01)***	28.919	5.212 (36.71)***	183.474	3.195 (28.44)***	24.420
FINSC	−0.083 (32.36)***	0.920	−0.084 (34.21)***	0.919	−0.074 (29.87)***	0.929	−0.075 (31.84)***	0.927
DMAJOR	0.272 (0.22)	1.313	0.234 (0.16)	1.264	−0.353 (0.56)	0.703	−0.382 (0.38)	0.683
DPULL	−2.124 (8.76)***	0.120			−2.330 (9.59)***	0.097		
DMIXED	−1.721 (6.69)**	0.179			−1.919 (7.03)***	0.147		
DPUSH			1.895 (10.00)***	6.654			2.110 (10.66)***	8.245
DLANGUAGE	−1.106 (2.85)*	0.331	−1.125 (2.93)*	0.325	−0.504 (0.44)	0.604	−0.516 (0.59)	0.597
Cox and Shell R^2	0.378		0.377		0.357		0.356	
Nagelkerke R^2	0.534		0.533		0.521		0.519	
HL χ^2 test	4.224		5.190		7.281		5.446	
Omnibus χ^2 test	82.135***		81.786***		76.435***		76.093***	

Source: author's estimations

Note: The Wald-statistics-statistics reported in parentheses are robust to heteroscedasticity. Letters in the superscript indicate significance levels: ***: 1 percent, **: 5 percent, *: 10 percent. P-values without a superscript suggest that the coefficient is insignificant even at the 10 percent level. HL: Hosmer and Lemeshow χ^2 test

4.2 Testing the Cognitive Bias of Self-assessment

A total of 120 students demonstrated an overestimation of their spreadsheet knowledge, while 53 students exhibited an underestimation compared to their results in the paper-based test. Similarly, compared to the Excel-based results, 127 students demonstrated an overestimation, while 46 students exhibited an underestimation of their spreadsheet knowledge. However, we ascertain the correlation between students' test results and the precision with which they over- or underestimate their proficiency in spreadsheet knowledge. A binary logistic regression model may prove the assumptions as an effective analytical tool. The dependent variable was measured on a dichotomous scale, with values of 1 indicating an overestimation and 0 indicating an underestimation.

The binary logistic regression analysis was conducted in Table 4 to ascertain the influence of students' self-assessments and their scores on the test. The analysis represents the proficiency results for the likelihood of participants' self-assessment, test results, majors, and digital learning methods. The Hosmer and Lemeshow (HL test) of the goodness of fit test suggests the models are found to be a good fit to the data as $p > 0.01$. The explained variation in the dependent variable based on models ranges from 35.6% to 37.8% (Cox and Shell R^2). The models explained 51.9–53.4% of the variance in self-assessment (Nagelkerke R^2). The overall significance tests of the models (Omnibus Tests of Model Coefficients), the coefficients and the odds ratios (Exp(β)) are reported in Table 3. The χ^2 tests ($df = 4$) determine a significant relationship between the independent variables and the binary dependent variable in both models.

The DK effect, which focuses directly on self-assessment relating to the extent of cognitive biases, is not independent of their sign. Individuals with low competence in a specific domain often exhibit overestimation of their abilities, while those with high competence may demonstrate underestimation. This phenomenon represents a form of biased self-assessment. The supposition is that lower-achieving students tend to overestimate their spreadsheet knowledge compared to the higher-achieving participants. Subsequently, it was determined that with each unit increase in test scores, irrespective of whether on paper or in Excel, there is a negative association between the Sprego knowledge estimates and the test scores. Consequently, higher-achieving students are more likely to underestimate their Sprego knowledge than the lower-achieving ones.

Furthermore, no correlation was identified between the self-assessment and the educational level (major) of the students. Concurrently, it was determined that students who underwent Sprego programming instruction tend to underestimate their spreadsheet knowledge, in comparison to their counterparts who still need to gain such experience. The students needing more familiarity with Sprego programming (PUSH) are inclined to overestimate their spreadsheet knowledge, in contrast to students who possess prior knowledge of Sprego. It was found that students without prior experience with Sprego were 6.654 times more likely to overestimate their spreadsheet knowledge in the paper test and 8.245 times more likely in the Excel test than those students who had previously learned Sprego programming. In both cases, DPULL and DMIXED regressors are negatively associated with the dependent variable, indicating that students with Sprego programming experience tend to underestimate their spreadsheet knowledge relative to students without such experience. Moreover, the findings revealed that in the comparison of language, international students exhibited less overestimation of spreadsheet knowledge in the case of paper tests compared to Hungarian students.

In order to elucidate the discrepancies amongst the cohorts, further methodologies are required to examine the manner in which the students appraise their spreadsheet knowledge. Educational background, nationality, learning methods and self-assessment disparities stratify the samples. Consequently, Mann–Whitney U-tests (Wilcoxon, 1945) are employed on the pre-assessment to ascertain which groups of learners tend to overestimate their knowledge in comparison to the total test score. The Mann–Whitney U-test represents a nonparametric

Table 5 Statistical characteristics of the pre-estimated and actual (paper and Excel) test scores

Cohorst	Scores	Mean Rank	Mean Rank Diff (MRD)	Mann–Whitney U-test	Z-value	P-value[a]
DMAJOR (n = 69)	Student	103.26	27.05	2466.0	−3.497	< 0.001***
	Paper	83.83	−5.27	3369.5	−0.678	0.498
	Excel	91.84	8.05	3254.0	−1.036	0.300
DLANGUAGE (n = 36)	Student	97.69	13.50	2081.0	−1.447	0.148
	Paper	64.31	−28.65	1649.0	−3.056	0.002***
	Excel	77.88	−11.68	2137.5	−1.229	0.219
DPULL (n = 27)	Student	65.59	−35.47	1393.0	−2.430	0.015**
	Paper	123.94	43.77	973.5	−4.173	< 0.001***
	Excel	115.63	33.98	1198.0	−3.235	0.001***
DMIXED (n = 39)	Student	75.27	−25.14	2155.5	−1.671	0.095*
	Paper	84.77	−2.98	2526.0	−0.316	0.752
	Excel	72.21	−19.10	2036.0	−2.097	0.036**
DPUSH (n = 107)	Student	96.68	25.37	2495.5	−3.253	0.001***
	Paper	78.49	−22.31	2620.5	−2.846	0.004***
	Excel	85.17	−4.80	3335.0	−0.613	0.540
OPREDIF (Paper) (n = 120)	Student	102.47	50.59	1324.0	−6.144	< 0.001***
	Paper	68.49	−60.43	958.5	−7.317	< 0.001***
	Excel	75.97	−35.98	1857.0	−4.358	< 0.001***
OPREDIF (Excel) (n = 127)	Student	99.55	47.20	1327.0	−5.506	< 0.001***
	Paper	75.67	−42.60	1482.5	−4.944	< 0.001***
	Excel	70.96	−61.69	884.5	−7.000	< 0.001***

Source: author's estimations

Note: N = 173. Significance of p-levels: ***: 1 percent, **: 5 percent, *: 10 percent. a = asymptotic 2-tailed p-values

statistical examination utilized to evaluate the hypothesis that two independent samples are derived from the same population or exhibit disparate population medians. The alternative hypothesis posits that the distributions are not identical. The discrepancies between the students' self-assessments and test scores can be summarized (see Table 5).

The results of the U-tests suggest significant differences between the two groups. Prior to the examination, the master students appeared to overestimate (MRD = 27.05) their proficiency in spreadsheets. The rejection of H2 is confirmed. Furthermore, our findings revealed that the mean ranks of students' paper-test scores exhibited notable variation by language, as indicated by the U-probes. The performance of international students is markedly inferior to that of Hungarian students (MRD = −28.65) in paper-based tests.

The results demonstrated that the mean rank scores of students on paper-based and Excel tests exhibited notable discrepancies across learning methods, as evidenced by the significant U-tests. The performance of students with no Sprego experience (DPUSH) on the tests was significantly lower than those with experience

(DPULL) in paper-based tests (MRD = 43.77 and Excel-tests (MRD = 33.98). Interestingly, those who use the mixed (DMIXED) learning methodology perform less well on the Excel test than the others (MRD = − 19.10). Consequently, based on the mean rank differences, we found that students with Sprego programming experience (DPULL) and (DMIXED) seem to predict spreadsheet knowledge more accurately than students without Sprego experience (DPUSH), MRD = − 35.47, MRD = − 25.14 and MRD = 25.37. Hypothesis H3 is confirmed. The findings display that the programming approach is more effective than the traditional tool-centered and mixed approach, even for students who were first exposed to the pull digital system during their tertiary studies. The failure of the mixed approach requires effort and assistance in transitioning to the pull system effectively and efficiently. Finally, we found that students who overestimate their performance compared to the paper-based and Excel test scores are notably different from their peers. They seem to perform significantly worse than those who underestimate their performance (MRD = − 60.43 and − 42.60 in Paper; MRD = − 35.98 and − 61.90 in Excel).

5 Discussion and Implications

The primary aim of this study was to evaluate the spreadsheet proficiency and self-evaluation of higher education students. A quantitative analysis was employed to assess the impact of educational level (majors) and spreadsheet programming skills on students' self-evaluation performance through traditional written and digital Excel tests.

The results of this study are consistent with previous research on cognitive biases and self-assessment in educational settings (Sawler, 2021; Surdilović et al., 2022). For example, investigations into the Dunning-Kruger effect, a well-documented cognitive bias, have demonstrated that students with lower achievement levels often overestimate their abilities (Kun, 2016). Conversely, high-achieving students tend to provide more accurate self-assessments and even underestimate their knowledge and competencies. Furthermore, our findings are in line with research on cognitive load and how fast and slow thinking can be effectively activated (Kahneman, 2011), indicating that students with greater expertise in CS-focused end-user areas (Sweller et al., 2011), such as Sprego programming in this case, tend to exhibit more precise self-assessment abilities. The findings align with the argument that spreadsheeting should be considered a form of programming and, as such, support building up fundamental CS knowledge.

Additionally, we assert that a spreadsheet is a more suitable form of programming for end-users when the emphasis is on the simplified functional language – including non-professionals in informatics and professionals in end-user roles (Csernoch et al., 2015) – compared to high-level programming languages (Sarkar et al., 2020). The common assumption that more experienced (higher-level) students possess superior self-assessment abilities due to their extensive academic background (Nepal & Kafle, 2024) is contradicted by this study. Despite the anticipation that students with certain majors would have a stronger grasp of spreadsheet skills, they demonstrated a tendency to overestimate their proficiency.

The primary novelty of this paper is its extensive research methods, which surpass the mere identification of cognitive biases to provide a profound understanding of self-assessment processes within the realm of Sprego programming. In contrast to prior studies that may have only superficially explored self-evaluation, this research delves into the complexities of how different factors – such as educational attainment, academic discipline, and Excel experience –interplay to impact students' self-assessment accuracy and, consequently, their awareness of errors and openness to avoid errors, reduce and eliminate waste.

The findings highlight the need for educational policies and practices that promote the development of accurate self-assessment skills along with science-based knowledge with an openness to continuous improvement (Chen et al., 2015), especially in areas that require technical expertise and problem-solving where these tools might be misleading (Wolfram, 2020). By addressing the cognitive biases that affect self-evaluation, educators can better prepare students for academic success and lifelong learning (Nierenberg & Dahl, 2023). This research not only provides valuable insights for educational theories but also offers practical implications for enhancing students' performance in diverse academic and cultural settings. By fostering improved self-assessment skills, educators can improve how students identify their strengths, weaknesses, and areas for development, leading to more problem-focused learning strategies and, ultimately, a higher level of proficiency.

We should consider how we can introduce spreadsheets into schools to develop computational thinking skills and fundamental CS knowledge. It is declared that developing computational thinking skills along with the 3Rs (Reading, wRiting, and aRithmetic) from an early mature is essential (Wing, 2006). Delaying this introduction reduces the likelihood of students developing computational thinking skills. Considering these factors, introducing CS- and programming-oriented spreadsheeting in primary and elementary schools can provide students with a solid foundation in CS knowledge (Lodi & Martini, 2021) that will remain relevant regardless of technological advancements, serving as a springboard for further studies.

One of the limitations of the study is that the sample needs to be more representative. Currently, the sample only includes students engaged in STEM-oriented tertiary education. Future research should prioritize the investigation of additional demographic segments, considering variables such as age, educational background, and professional inclinations. Previous research has shown that Sprego programming can be effectively introduced from Grade 7 onwards in various groups and orientations (Csernoch et al., 2015; Csapó et al., 2020, 2021; Takács and Bubnó, 2022).

Moreover, there was no significant variance in the academic performance of students when comparing the use of paper versus Excel. However, it is worth noting that the time allocated for utilizing these two different methods varied. In this initial test, students were given unrestricted time to complete the task, and our findings indicated that students required two to three times more time in Excel than on paper. In order to ensure more accurate and standardized outcomes, it is essential to record students' activities simultaneously with the completion of the associated paperwork. One of our future challenges is to ascertain their progress simultaneously at the same time.

The cognitive biases uncovered in this study related to self-assessment accuracy are not limited to computer science. They are likely applicable across various scientific disciplines, especially those that necessitate intricate problem-solving abilities and technical expertise, and in cases when various school subjects are taught with digital support. In these fields, students often need to assess their knowledge and abilities to succeed accurately by integrating subject and technological knowledge, skills, and abilities. Conversely, underestimation could lead to unnecessary anxiety or reluctance to tackle challenging problems. Recognizing these cognitive biases could aid educators in devising strategies to enhance self-assessment accuracy across sciences, resulting in improved learning outcomes and more effective teaching practices. Many countries encounter similar educational challenges as those observed in Hungary with comparable educational structures and student demographics.

This research makes a significant contribution to the field by deepening the understanding of how cognitive biases affect self-evaluation in the context of spreadsheet problem-solving and beyond within the broader higher educational landscape. The study unveils the intricate relations between variables such as educational background, academic major, and internalization and how these factors collectively influence students' ability to assess their skills and knowledge accurately.

6 Conclusion

This research paper sheds light on the current state of biased self-assessment in spreadsheet programming. By seeking the relevant literature, we identified the most influential aspects that can influence how students self-evaluate their knowledge. Moreover, the regression and non-parametric statistical analyses provided a rigorous perspective on the relationships between different concepts in the literature and highlighted areas for further research.

The research has pinpointed numerous crucial elements that influence cognitive biases in end-user experience. According to the proposed framework, the outcomes indicate that (1) higher-achieving students reveal better ability to evaluate their exam results compared to their lower-achieving peers, and they are more likely to underestimate their spreadsheet knowledge than the lower-achieving ones. (2) The master students are less likely to be accurate in their proficiency in spreadsheet usage compared to their undergraduate counterparts. Findings indicate that even higher education does not necessarily equate to better self-awareness in specific skill areas. (3) The students with experience in Sprego programming appear to be more accurate in predicting their spreadsheet proficiency instruction and tend to underestimate their spreadsheet knowledge in comparison to their counterparts without Sprego experience. Interestingly, we found that students who overestimate their performance compared to the paper-based and Excel test scores are significantly different from their peers. They perform worse than those who underestimate their performance.

In sum, the findings present a unified account that links specific cognitive biases, such as overconfidence and underestimation, to academic performance in the context of spreadsheet programming. Students who perform well and have experience with

Sprego programming tend to have more accurate self-assessments, although they may be inclined to underestimate their abilities. Conversely, master students and those who overestimate their abilities often face the challenge of misjudging their proficiency, which correlates with lower actual performance. It is important to increase the representation of students at the bachelor and master levels in areas that are under-represented in higher education, such as science and engineering, and to reduce students' drop-out rates. This study underscores the significance of metacognitive awareness in higher education, particularly in programming-related subjects, where self-assessment accuracy can significantly impact learning outcomes. The research reveals a notable difference between students who overestimate their performance and those who underestimate it. The former group performed significantly worse in both paper-based and Excel tests, indicating that overconfidence leads to poorer outcomes by hindering students from recognizing and addressing knowledge gaps.

However, there are also challenges associated with the results, including the need for ongoing support beyond the findings. The results imply that Hungarian higher education can improve student academic success and lifelong learning by tackling cognitive biases and bolstering metacognitive skills. This approach could result in a more adept and flexible workforce in the future. Hence, it would be beneficial to develop a national framework for the teaching of metacognitive skills, integrate Sprego programming into learning programs, provide targeted support for different groups of students, review assessment methods and promote a culture of continuous learning and accurate self-evaluation.

In conclusion, this study not only deepens the understanding of the cognitive factors that influence student performance in programming-related spreadsheet tasks but also highlights critical areas for improvement in higher education. By addressing these cognitive biases through informed policy decisions, Hungarian higher education can better equip students with the skills and awareness needed for academic and business success. The importance of these findings lies in their potential to shape a more effective and responsive educational system, ultimately fostering a generation of students who are both confident and competent in their abilities.

The study has extended limitations. (1) The limited diversity of the sample population might constrain the findings. Hence, the results may not be easily generalized to the broader student population in Hungary or other educational settings. A more varied sample, including students from different educational institutions, fields of study, and academic levels, could provide a more comprehensive understanding of cognitive biases in spreadsheet programming. (2) Factors such as the quality of teaching, access to resources, and the learning environment also significantly affect self-assessment accuracy. The lack of a thorough examination of these external influences could limit our understanding of why certain groups, such as master students, tend to overestimate their abilities and how these factors interact with cognitive biases. (3) Self-reported measures may not always accurately reflect a student's true proficiency or cognitive biases, as factors such as social desirability, memory recall, and personal perceptions can influence them. Relying on self-reported data could limit the accuracy of the findings. Future research could incorporate more objective measures of student performance and cognitive biases, such as observational data or controlled experiments.

Abbreviations UX: User Experience; CS: Computer Science; CER: Computer Education Research; DK: Dunning-Kruger; STEM: Science, Technology, Engineering and Mathematics; SPREGO: Spreadsheet Lego; TPS: Toyota Production System; MSc: Bachelor of Science (BSc) and Master of Science; MRD: Mean Rank Difference; 3Rs: Reading, wRiting, and aRithmetic

Acknowledgements Supported by the University of Debrecen Program for Scientific Publication.

Authors' contributions All authors listed in this manuscript, made substantial contributions to the conception, design, testing process, and writing of this work, including drafting the work and critical revisions.

Funding Open access funding provided by University of Debrecen.

Data availability The data that support the findings of this study are available from the corresponding author upon request. The authors confirm that all data generated or analyzed during this study are included in this published article.

Code availability Not applicable.

Declarations

Ethics approval Not applicable.

Consent to participate Not applicable.

Consent for publication All authors have fully agreed to this submission and agree to be accountable for all aspects of the work in ensuring that questions related to the accuracy or the integrity of any part of the work is appropriately investigated and resolved.

Conflict of interest The authors declared no potential conflict of interest with respect to the authorship and/or publication of this article.

References

Abraham, R., Burnett, M., & Erwig, M. (2009). Spreadsheet Programming. In *Wiley Encyclopedia of Computer Science and Engineering* (pp. 2804–2810). Wiley. https://doi.org/10.1002/9780470050118.ecse415

Alicke, M. D., & Sedikides, C. (2009). Self-enhancement and self-protection: What they are and what they do. *European Review of Social Psychology, 20*(1), 1–48. https://doi.org/10.1080/10463280802613866

Anson, I. G. (2018). Partisanship, Political Knowledge, and the Dunning-Kruger Effect. *Political Psychology, 39*(5), 1173–1192. https://doi.org/10.1111/POPS.12490

Azzopardi, L. (2021). Cognitive Biases in Search: A Review and Reflection of Cognitive Biases in Information Retrieval. *CHIIR 2021 - Proceedings of the 2021 Conference on Human Information Interaction and Retrieval*, 27–37. https://doi.org/10.1145/3406522.3446023

Balasubramnian, B., & Sargent, C. S. (2020). Impact of inflated perceptions of financial literacy on financial decision making. *Journal of Economic Psychology, 80*, 102306. https://doi.org/10.1016/J.JOEP.2020.102306

Baranyi, P., & Gilanyi, A. (2013). Mathability: Emulating and enhancing human mathematical capabilities. In *2013 IEEE 4th International Conference on Cognitive Infocommunications (CogInfoCom)* (pp. 555–558). IEEE. https://doi.org/10.1109/CogInfoCom.2013.6719309

Blanco, F. (2017). Cognitive Bias. *Encyclopedia of Animal Cognition and Behavior*, 1–7. https://doi.org/10.1007/978-3-319-47829-6_1244-1

Booth, Shirley. (1992). *Learning to program : a phenomenographic perspective*. Acta Universitatis Gothoburgensis. Accessed 15 August 2024

Burnett, M., Atwood, J., Djang, R. W., Reichwein, J., Gottfried, H., & Yang, S. (2001). Forms/3: A first-order visual language to explore the boundaries of the spreadsheet paradigm. *Journal of Functional Programming*, 155–206. https://doi.org/10.5555/968486.968487

Carroll, J. M. . (1987). *Interfacing thought : cognitive aspects of human-computer interaction*. MIT Press. Accessed 23 December 2024

Chen, J., Morris, D., & Mansour, N. (2015). Science teachers' beliefs: Perceptions of efficacy and the nature of scientific knowledge and knowing. In H. Fives, & M. G. Gill (Eds.), *International Handbook of Research on Teachers' Beliefs* (pp. 370–386). Routledge.

Collins, J. (2001). *Good to Great: Why Some Companies Make the Leap … and Some Others Don't*. Random House Business.

Crawford, J. D., & Stankov, L. (1996). Age differences in the realism of confidence judgements: A calibration study using tests of fluid and crystallized intelligence. *Learning and Individual Differences, 8*(2), 83–103. https://doi.org/10.1016/S1041-6080(96)90027-8

Csapó, G., Csernoch, M., & Abari, K. (2020). Sprego: Case study on the effectiveness of teaching spreadsheet management with schema construction. *Education and Information Technologies, 25*(3), 1585–1605. https://doi.org/10.1007/S10639-019-10024-2/TABLES/8

Csapó, G., Sebestyén, K., Csernoch, M., & Abari, K. (2021). Case study: Developing long-term knowledge with Sprego. *Education and Information Technologies, 26*(1), 965–982. https://doi.org/10.1007/S10639-020-10295-0/TABLES/7

Csernoch, M. (2014). *Programozás táblázatkezelő függvényekkel*. Műszaki Tankkönyvkiadó.

Csernoch, M., & Bíró, P. (2015). Problem solving in Sprego. In *Proc. 16th EuSpRIG Conf. Spreadsheet Risk Management* (pp. 1–13).

Csernoch, M., Biró, P., Máth, J., & Abari, K. (2015). Testing Algorithmic Skills in Traditional and Non-Traditional Programming Environments. *Informatics in Education., 14*(2), 175–197. https://doi.org/10.15388/INFEDU.2015.11

Csernoch, M., Szűcs, E. G., & Máté, D. (2024a). Kanban method in digital data processing. *International Review of Applied Sciences and Engineering*, (Online first). https://doi.org/10.1556/1848.2024.00909

Csernoch, M., T. Kiss, J., Takács, V. L., & Máté, D. (2024b). Exploring Higher Education Competencies through Spreadsheet Self-Assessment & Time. In *European Spreadsheet Risks Interest Group: Spreadsheet Productivity & Risks* (pp. 47–62). https://m2.mtmt.hu/api/publication/35209542

de Bruin, A. B. H., Kok, E. M., 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 and Learning, 12*(1), 21–43. https://doi.org/10.1007/S11409-016-9159-5/TABLES/8

Dunning, D. (2011). The Dunning-Kruger Effect: On Being Ignorant of One's Own Ignorance. *Advances in Experimental Social Psychology, 44*, 247–296. https://doi.org/10.1016/B978-0-12-385522-0.00005-6

Dunning, D., Johnson, K., Ehrlinger, J., & Kruger, J. (2003). Why People Fail to Recognize Their Own Incompetence. *Current Directions in Psychological Science, 12*(3), 83–87. https://doi.org/10.1111/1467-8721.01235

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

Eva, K. W., Regehr, G., & Gruppen, L. D. (2018). 6. Blinded By "Insight." In B. D. Hodges & L. Lingard (Eds.), *The Question of Competence* (pp. 131–154). Ithaca, NY: Cornell University Press. https://doi.org/10.7591/9780801465802-010

Feld, J., Sauermann, J., & de Grip, A. (2017). Estimating the relationship between skill and overconfidence. *Journal of Behavioral and Experimental Economics, 68*, 18–24. https://doi.org/10.1016/J.SOCEC.2017.03.002

Frank, P., Henkel, G., & Lysgaard, J. A. (2024). Between evidence and delusion – a scoping review of cognitive biases in Environmental and Sustainability Education. *Environmental Education Research, 30*(9), 1477–1499. https://doi.org/10.1080/13504622.2024.2371507

Friedman, H. H. (2023). Cognitive Biases and Their Influence on Critical Thinking and Scientific Reasoning: A Practical Guide for Students and Teachers. *SSRN Electronic Journal*. https://doi.org/10.2139/SSRN.2958800

Gibbs, S., Moore, K., Steel, G., & McKinnon, A. (2017). The Dunning-Kruger Effect in a workplace computing setting. *Computers in Human Behavior, 72*, 589–595. https://doi.org/10.1016/J.CHB.2016.12.084

Gignac, G. E. (2022). The association between objective and subjective financial literacy: Failure to observe the Dunning-Kruger effect. *Personality and Individual Differences, 184*, 111224. https://doi.org/10.1016/J.PAID.2021.111224

Gignac, G. E. (2024). Rethinking the Dunning-Kruger effect: Negligible influence on a limited segment of the population. *Intelligence, 104*, 101830. https://doi.org/10.1016/j.intell.2024.101830

Gonda, T. (2022). Observing the Dunning-Kruger Effect in the Results of Tourism Studies. *Journal of Environmental Management and Tourism, 13*(6), 1551–1556. https://doi.org/10.14505/JEMT.V13.6(62).03

Hair, J. F. (2010). *Multivariate data analysis*. Prentice Hall.

Han, J., Kelley, T., & Knowles, J. G. (2021). Factors Influencing Student STEM Learning: Self-Efficacy and Outcome Expectancy, 21st Century Skills, and Career Awareness. *Journal for STEM Education Research, 4*(2), 117–137. https://doi.org/10.1007/S41979-021-00053-3/TABLES/8

Handley, M. A., Lyles, C. R., McCulloch, C., & Cattamanchi, A. (2018). Selecting and Improving Quasi-Experimental Designs in Effectiveness and Implementation Research. *Annual Review of Public Health*, *39*(Volume 39, 2018), 5–25. https://doi.org/10.1146/ANNUREV-PUBLHEALTH-040617-014128/CITE/REFWORKS

Hartwig, M. K., & Dunlosky, J. (2014). The contribution of judgment scale to the unskilled-and-unaware phenomenon: How evaluating others can exaggerate over- (and under-) confidence. *Memory & Cognition, 42*(1), 164–173. https://doi.org/10.3758/S13421-013-0351-4

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

Hubwieser, P. (2004). Functional Modelling in Secondary Schools Using Spreadsheets. *Education and Information Technologies 2004 9:2*, *9*(2), 175–183. https://doi.org/10.1023/B:EAIT.0000027929.91773.AB

Jansen, R. A., Rafferty, A. N., & Griffiths, T. L. (2021). A rational model of the Dunning-Kruger effect supports insensitivity to evidence in low performers. *Nature Human Behaviour, 5*(6), 756–763. https://doi.org/10.1038/s41562-021-01057-0

Jia, Z., Li, S., Luo, Z., Tong, M., & Gao, T. (2023). The Dunning-Kruger effect: Subjective health perceptions on smoking behavior among older Chinese adults. *BMC Public Health, 23*(1), 1703. https://doi.org/10.1186/S12889-023-16582-Y

Jin, K. Y., Reichert, F., Cagasan, L. P., de la Torre, J., & Law, N. (2020). Measuring digital literacy across three age cohorts: Exploring test dimensionality and performance differences. *Computers & Education, 157*, 103968. https://doi.org/10.1016/J.COMPEDU.2020.103968

Kahneman, D. (2011). *Thinking*. Penguin.

Kahneman, Daniel., Sibony, Olivier., & Sunstein, C. R. (2021). *Noise a Flaw in Human Judgement*. London: William Collins. https://search.worldcat.org/title/1257248287. Accessed 23 December 2024

Kakinohana, R. K., & Pilati, R. (2023). Differences in decisions affected by cognitive biases: Examining human values, need for cognition, and numeracy. *Psicologia, Reflexão e Crítica : Revista Semestral Do Departamento De Psicologia Da UFRGS, 36*(1), 1–14. https://doi.org/10.1186/S41155-023-00265-Z

Karpen, S. C. (2018). The Social Psychology of Biased Self-Assessment. *American Journal of Pharmaceutical Education, 82*(5), 441–448. https://doi.org/10.5688/AJPE6299

Kim, Y. H., Kwon, H., Lee, J., & Chiu, C. Y. (2016). Why Do People Overestimate or Underestimate Their Abilities? *A Cross-Culturally Valid Model of Cognitive and Motivational Processes in Self-Assessment Biases*. https://doi.org/10.1177/0022022116661243

Kim, S., Raza, M., & Seidman, E. (2019). Improving 21st-century teaching skills: The key to effective 21st-century learners. *Research in Comparative and International Education, 14*(1), 99–117. https://doi.org/10.1177/1745499919829214

Kirschner, P. A., & De Bruyckere, P. (2017). The myths of the digital native and the multitasker. *Teaching and Teacher Education, 67*, 135–142. https://doi.org/10.1016/J.TATE.2017.06.001

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

Kun, A. I. (2016). A comparison of self versus tutor assessment among Hungarian undergraduate business students. *Assessment & Evaluation in Higher Education, 41*(3), 350–367. https://doi.org/10.1080/02602938.2015.1011602

Lebuda, I., Hofer, G., Rominger, C., & Benedek, M. (2024). No strong support for a Dunning-Kruger effect in creativity: Analyses of self-assessment in absolute and relative terms. *Scientific Reports, 14*(1), 11883. https://doi.org/10.1038/s41598-024-61042-1

Liker, J. (2004). *The Toyota Way: 14 Management Principles from the World's Greatest Manufacturer*. McGraw Hill.

Lister, R. (2008). After the Gold Rush: Toward Sustainable Scholarship in Computing. *ACE '08: Proceedings of the tenth conference on Australasian computing education*, (78), 3–17. https://doi.org/10.5555/1379249.1379269

Lodi, M., & Martini, S. (2021). Computational Thinking, Between Papert and Wing. *Science and Education, 30*(4), 883–908. https://doi.org/10.1007/S11191-021-00202-5/METRICS

Lu, J., Song, Y., Li, Q., Tang, J., & Hou, Y. (2023). Research on the Price Prediction of Bitcoin and Gold Based on Random Forest Model. In *Proceedings of the 2nd International Conference on Information, Control and Automation, ICICA 2022, December 2–4, 2022, Chongqing, China*. EAI. https://doi.org/10.4108/eai.2-12-2022.2327931

MacNeil, S. L., Wood, E., & Arslantas, F. (2024). Development of a metacognition co-curriculum for a university course in introductory organic chemistry. *Frontiers in Education*, *9*. https://doi.org/10.3389/feduc.2024.1402599

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

Mahmood, K. (2016). Do People Overestimate Their Information Literacy Skills? A Systematic Review of Empirical Evidence on the Dunning-Kruger Effect. *Comminfolit*, *10*(2), 199. https://doi.org/10.15760/comminfolit.2016.10.2.24

Malmi, L., Sheard, J., Kinnunen, P., Simon, & Sinclair, J. (2019). Computing education theories: what are they and how are they used? *ICER 2019 - Proceedings of the 2019 ACM Conference on International Computing Education Research*, 187–197. https://doi.org/10.1145/3291279.3339409

Malmi, L., Sheard, J., Kinnunen, P., Simon, & Sinclair, J. (2023). Development and Use of Domain-specific Learning Theories, Models, and Instruments in Computing Education. *ACM Transactions on Computing Education*, *23*(1), 1–48https://doi.org/10.1145/3530221

Máté, D., & Darabos, É. (2017). Measuring the Accuracy of Self-assessment Among Undergraduate Students in Higher Education to Enhance Competitiveness. *Journal of Competitiveness, 9*(2), 78–92. https://doi.org/10.7441/joc.2017.02.06

McIntosh, R. D., Fowler, E. A., Lyu, T., & Sala, S. D. (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

Moore, D., & Healy, P. (2008). The Trouble With Overconfidence. *Psychological Review, 115*, 502–517. https://doi.org/10.1037/0033-295X.115.2.502

Motta, M., Callaghan, T., & Sylvester, S. (2018). Knowing less but presuming more: Dunning-Kruger effects and the endorsement of anti-vaccine policy attitudes. *Social Science & Medicine, 1982*(211), 274–281. https://doi.org/10.1016/J.SOCSCIMED.2018.06.032

Nagy, K., & Csernoch, M. (2023). Pre-testing erroneous text-based documents: Logging end-user activities. *Frontiers in Education, 7*, 958635. https://doi.org/10.3389/FEDUC.2022.958635/BIBTEX

Nagy, T., Csernoch, M., & Biró, P. (2021). The Comparison of Students' Self-Assessment, Gender, and Programming-Oriented Spreadsheet Skills. *Education Sciences, 11*(10), 590. https://doi.org/10.3390/EDUCSCI11100590

Nepal, K., & Kafle, R. C. (2024). Who can predict their performance more accurately? An investigation of undergraduate students' self-assessment behavior in mathematics courses. *Metacognition and Learning, 19*(2), 549–565. https://doi.org/10.1007/S11409-024-09381-2/METRICS

Nierenberg, E., & Dahl, T. I. (2023). Is information literacy ability, and metacognition of that ability, related to interest, gender, or education level? A cross-sectional study of higher education students. *Journal of Librarianship and Information Science, 55*(1), 57–69. https://doi.org/10.1177/09610006211058907/ASSET/IMAGES/LARGE/10.1177_09610006211058907-FIG2.JPEG

Niklas, M., & Pär, Å. (2015). *This is Lean: Resolving the Efficiency Paradox*. Rheologica Publishing.

Ohno, T. (1988). Toyota Production System: Beyond Large-Scale Production. *Toyota Production System: Beyond Large-Scale Production. New York: Taylor and Francis.* https://doi.org/10.4324/9780429273018/TOYOTA-PRODUCTION-SYSTEM-TAIICHI-OHNO

Pak, T. Y., & Chatterjee, S. (2016). Aging, overconfidence, and portfolio choice. *Journal of Behavioral and Experimental Finance, 12*, 112–122. https://doi.org/10.1016/J.JBEF.2016.10.003

Panko, R. R. (2013). The Cognitive Science of Spreadsheet Errors: Why Thinking is Bad. In *2013 46th Hawaii International Conference on System Sciences* (pp. 4013–4022). IEEE. https://doi.org/10.1109/HICSS.2013.513

Polya, G. (1945). *How to solve it*. Princeton University Press.

Prims, J. P., & Moore, D. A. (2017). Overconfidence over the lifespan. *Judgment and Decision Making, 12*(1), 29–41. https://doi.org/10.1017/S1930297500005222

Rattenbury, T., Hellerstein, J. M., Heer, J., Kandel, S., & Carreras, C. (2017). *Principles of Data Wrangling: Practical Techniques for Data Preparation*. O'Reilly Media.

Reynolds, G. (2008). *Presentation Zen: Simple Ideas on Presentation Design and Delivery*. New Riders Pub.

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

Rother, M. (2003). *Learning to See Value Stream Mapping to Create Value and Eliminate Muda*. *Lean Enterprise Institute Brookline*. Lean Enterprise Institute. https://doi.org/10.1109/6.490058

Rother, M. (2009). *Toyota Kata: Managing People for Improvement, Adaptiveness and Superior Results*. McGraw Hill.

Sarkar, A., Borghouts, J., Iyer, A., Khullar, S., Canton, C., Hermans, F., et al. (2020). Spreadsheet Use and Programming Experience: An Exploratory Survey (pp. 1–9). https://doi.org/10.1145/3334480.3382807

Sawler, J. (2021). Economics 101-ism and the Dunning-Kruger effect: Reducing overconfidence among introductory macroeconomics students. *International Review of Economics Education, 36*, 100208. https://doi.org/10.1016/j.iree.2020.100208

Schaefer, S., Riediger, M., Li, S. C., & Lindenberger, U. (2023). Too easy, too hard, or just right: Lifespan age differences and gender effects on task difficulty choices. *International Journal of Behavioral Development, 47*(3), 253–264. https://doi.org/10.1177/01650254231160126

Scheiber, R., Karmasin, M., & Diehl, S. (2023). Exploring the Dunning-Kruger Effect in Health Communication: How Perceived Food and Media Literacy and Actual Knowledge Drift Apart When Evaluating Misleading Food Advertising. *Journal of Health Communication, 28*(11), 707–727. https://doi.org/10.1080/10810730.2023.2258085

Schunk, D., & Zimmerman, B. (1997). Social origin of self-regulatory competence. *Educational Psychologist, 32*, 195–208. https://doi.org/10.1207/s15326985ep3204_1

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

Surdilović, D., Adtani, P., Fuoad, S. A., Abdelaal, H. M., & D'souza, J. (2022). Evaluation of the Dunning-Kruger Effects among Dental Students at an Academic Training Institution in UAE. *Acta stomatologica Croatica, 56*(3), 299–310. https://doi.org/10.15644/ASC56/3/8

Sweller, J., Kirschner, P., & Clark, R. E. (2007). Why Minimally Guided Teaching Techniques Do Not Work: A Reply to Commentaries. *Educational Psychologist, 42*(2), 115–121. https://doi.org/10.1080/00461520701263426

Sweller, J., Ayres, P., & Kalyuga, S. (2011). Cognitive Load Theory in Perspective. In *Cognitive Load Theory* (pp. 237–242). New York, NY: Springer New York. https://doi.org/10.1007/978-1-4419-8126-4_18

Turner, P. (2020). Critical values for the Durbin-Watson test in large samples. *Applied Economics Letters, 27*(18), 1495–1499. https://doi.org/10.1080/13504851.2019.1691711 von Suchodoletz, A., & Hepach, R. (2021). Cultural values shape the expression of self-evaluative social emotions. *Scientific Reports, 11*(1), 1–14. https://doi.org/10.1038/s41598-021-92652-8

Wakeling, D. (2007). Spreadsheet functional programming. *Journal of Functional Programming, 17*(1), 131–143. https://doi.org/10.1017/S0956796806006186

Wilcoxon, F. (1945). Individual Comparisons by Ranking Methods. *Biometrics Bulletin, 1*(6), 80–83. https://doi.org/10.2307/3001968

Wing, J. M. (2006). Computational thinking. *Communications of the ACM, 49*(3), 33–35. https://doi.org/10.1145/1118178.1118215

Wolfram, C. (2020). *The math(s) fix: an education blueprint for the AI age*. Wolfram Media, Inc. Accessed 16 August 2024

Wolfram, C. (2024). Evidence: let's promote not stifle innovation in education — Conrad Wolfram. https://www.conradwolfram.com/writings/2015/5/21/role-of-evidence-in-education-innovation. Accessed 4 September 2024. Accessed 4 September 2024

Womack, J. P., & Jones, D. T. (1997). Lean Thinking—Banish Waste and Create Wealth in your Corporation. *Journal of the Operational Research Society, 48*(11), 1148–1148. https://doi.org/10.1038/sj.jors.2600967

Xia, M., Poorthuis, A. M. G., & Thomaes, S. (2024). Children's overestimation of performance across age, task, and historical time: A meta-analysis. *Child Development, 95*(3), 1001–1022. https://doi.org/10.1111/CDEV.14042

Yang, H., Thorsen, C. H., Radišić, J., Peixoto, F., Laine, A., & Liu, X. (2024). When competence and confidence are at odds: A cross-country examination of the Dunning-Kruger effect. *European Journal of Psychology of Education, 39*(2), 1537–1559. https://doi.org/10.1007/S10212-024-00804-X/TABLES/7

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Authors and Affiliations

Domicián Máté[1,2] · **Judit T. Kiss**[1] · **Mária Csernoch**[3]

✉ Judit T. Kiss
tkiss@eng.unideb.hu

Domicián Máté
mate.domician@eng.unideb.hu

Mária Csernoch
csernoch.maria@inf.unideb.hu

[1] Department of Engineering Management and Enterprise, Faculty of Engineering, University of Debrecen, Debrecen, Hungary

[2] DHET-NRF Sarchi Entrepreneurship Education, Department of Business Management, University of Johannesburg, Johannesburg, South Africa

[3] Department of Computer Science, Faculty of Informatics, University of Debrecen, Debrecen, Hungary