Which One Is the “Best”: a Cross-national Comparative Study of Students’ Strategy Evaluation in Equation Solving

Citation for the original published paper (version of record):
Jiang, R., Star, J., Hästö, P. et al (2022). Which One Is the “Best”: a Cross-national Comparative Study of Students’ Strategy Evaluation in
Equation Solving. International Journal of Science and Mathematics Education, In press.
http://dx.doi.org/10.1007/s10763-022-10282-6

International Journal of Science and Mathematics Education
https://doi.org/10.1007/s10763-022-10282-6

Which One Is the "Best": a Cross-national Comparative Study of Students' Strategy Evaluation in Equation Solving

Ronghuan Jiang[1] · Jon R. Star[2] · Peter Hästö[3] · Lijia Li[2] · Ru-de Liu[4] · Dimitri Tuomela[3] · Nuria Joglar Prieto[5] · Riikka Palkki[3] · Miguel Á. Abánades[6] · Johanna Pejlare[7]

Received: 19 August 2021 / Accepted: 5 April 2022

Abstract

This cross-national study examined students' evaluation of strategies for solving linear equations, as well as the extent to which their evaluation criteria were related to their use of strategies and/or aligned with experts' views about which strategy is the best. A total of 792 middle school and high school students from Sweden, Finland, and Spain participated in the study. Students were asked to solve twelve equations, provide multiple solving strategies for each equation, and select the best strategy among those they produced for each equation. Our results indicate that students' evaluation of strategies was not strongly related to their initial preferences for using strategies. Instead, many students' criteria were aligned with the flexibility goals, in that a strategy that takes advantages of task context was more highly valued than a standard algorithm. However, cross-national differences in strategy evaluation indicated that Swedish and Finnish students were more aligned with flexibility goals in terms of their strategy evaluation criteria, while Spanish students tended to consider standard algorithms better than other strategies. We also found that high school students showed more flexibility concerns than middle school students. Different emphases in educational practice and prior knowledge might explain these cross-national differences as well as the findings of developmental changes in students' evaluation criteria.

Keywords Strategy evaluation · Cross-national study · Flexibility · Equation solving · Algebra

✉ Ru-de Liu
rdliu@bnu.edu.cn

Extended author information available on the last page of the article

Published online: 16 May 2022

Introduction

Algebra is considered a gateway to high-level mathematics and future learning opportunities such as entrance into college (Impecoven-Lind & Foegen, 2010). For the past decades, researchers have investigated how to improve students' algebraic skills, with one stream of research focused on developing students' flexibility in algebra problem-solving (Hatisaru, 2021; Rittle-Johnson & Star, 2007; Schneider et al., 2011; Star & Newton, 2009; Xu et al., 2017). Within this literature, flexibility is defined as the ability to generate, use, and evaluate multiple strategies for a given problem. As such, it has been emphasized in many policy documents across the world (Australian Education Ministers, 2006; National Mathematics Advisory Panel, 2008; Singapore Ministry of Education, 2006; Woodward et al., 2012). Perhaps as a result, in the domain of algebraic problem-solving, studies have focused on measuring strategy flexibility (Newton et al., 2019; Star & Seifert, 2006; Xu et al., 2017), understanding important factors that could affect strategy flexibility (DeCaro, 2016; Jiang et al., 2021; Keleş & Yazgan, 2021; Ramirez et al., 2016; Shaw et al., 2020; Threlfall, 2009; Wang et al., 2019), and how instructional interventions could facilitate strategy flexibility (De Smedt et al., 2010; Nistal et al., 2014; Star & Rittle-Johnson, 2008; Star et al., 2015a, b).

In most of these studies, one important aspect of flexibility is the ability to evaluate the most appropriate strategy for a given problem. Determining the relative merits of various strategies for solving problems is central to the work of doing mathematics (Carroll, 1999; Star & Madnani, 2004; Verschaffel et al., 2009). The ability to justify the advantages and disadvantages of problem-solving strategies is crucial for strategy use and selection (Luwel et al., 2003). Thus, it is important to understand how students justify the appropriateness of a strategy. Yet it is also the case that what it means for a strategy to be appropriate is often dependent on the solver's beliefs and goals as well as the particular tasks to be solved (Verschaffel et al., 2009). However, few studies have explicitly investigated how students select and evaluate their strategies and how different factors could affect their evaluation of strategy appropriateness. Thus, the present study aimed at exploring students' strategy evaluation criteria they used when determining which strategies were considered appropriate in equation solving. In addition, we applied a cross-national design to further understand how different educational settings may affect students' strategy use and strategy evaluation.

Evaluating the Appropriateness of strategies: The Importance of the Task Context

Prior studies that emphasized the development of adaptive expertise (Baroody, 2003; Baroody & Dowker, 2013; Verschaffel et al., 2009) suggested that the ability to solve mathematical problems efficiently and flexibly with different strategies is crucial for mathematical problem-solving. In contrast with routine expertise (i.e. solve mathematical problems by applying algorithms), adaptive expertise allows students to solve problems in meaningful ways and to select the most appropriate

Table 1 Sample equation and solution methods from Xu et al. (2017)

Sample equation	Strategy 1	Strategy 2
$3(x-1)=27$	$3x-3=27$ $3x=30$ $x=10$	$x-1=9$ $x=10$

strategy regarding the characteristics of the mathematical task (Baroody & Dowker, 2013; Verschaffel et al., 2009). These researchers suggested that an adaptive expert should be able to evaluate different strategies and use the strategy that is appropriate according to the specific features of the problem at hand. In this literature, the appropriateness of a strategy is mainly depending on the task contexts.

Evaluating the appropriateness or merits of the strategy based on the task context is also an important part of a learner's meta-strategic knowledge (Karlen, 2015; Kuhn & Pearsall, 1998; Zohar, 2012; Zohar & David, 2008; Zohar & Peled, 2008). Meta-strategic knowledge, which is a type of explicit higher-order thinking, has two distinct but related components: knowledge of tasks and knowledge of strategies. Together, these two types of meta-strategic knowledge allow students to think about the objectives of a given task; the task characteristics that suggest the use of a specific strategy; when, why, and how a strategy should be used; and the disadvantages of not using appropriate strategies (Zohar, 2012; Zohar & David, 2008). In terms of understanding the task, meta-strategic knowledge implies that strategies are applied in the service of the goals and purposes of particular tasks (Zohar & Peled, 2008). Thus, a learner needs to have a good understanding of the task—particularly meta-strategic knowledge about the task—in order to select appropriate strategies. Researchers have found that meta-strategic knowledge is a key factor that could influence students' strategy performance (Luwel et al., 2003) and the evaluation of strategy use (Karlen, 2015).

The appropriateness of a strategy has also been found to be related to the task context in prior studies that focused on strategic flexibility in equation solving. Students' knowledge of appropriate strategies has generally been measured by a two-part process. Students were first asked to solve several problems and to provide as many strategies as they could. Students were then asked to select the strategy that they considered to be the best (Liu et al., 2018; Star et al., 2015a, b; Wang et al., 2019; Xu et al., 2017). These studies highlighted innovativeness and efficiency as the defining criteria for selecting appropriate strategies.

For example, according to Xu et al. (2017), if a student solved the equation in Table 1 using the two strategies shown and then identified Strategy 2 as the best one, then the student was recognized as having chosen the appropriate strategy for this problem. Specifically, in Xu et al. (2017) (see also Star & Seifert, 2006; Star et al., 2022; Rittle-Johnson & Star, 2007), Strategy 2 is considered a more appropriate strategy in this particular task context because it reduces computational steps (Star & Seifert, 2006). A flexible problem solver should be able to solve equations such as the one shown in Table 1 using multiple ways and to identify that Strategy 2 is more appropriate (Liu et al., 2018; Xu et al., 2017). Drawing from this perspective, we refer to strategies that might be considered more appropriate or better than other strategies based on task contexts as “situationally appropriate.” The learning of situationally appropriate strategies is linked in prior research to flexibility considerations (Star & Rittle-Johnson, 2008).

The use of evaluation criteria for strategy appropriateness based on task context is also seen among mathematics experts. In particular, Star and Newton (2009) interviewed eight experts including mathematicians, mathematics teachers, and engineers. They asked these experts to complete a mathematics test which included several linear equations that were similar to the example presented in Table 1. After the experts completed the test, researchers asked them to determine which strategy or strategies that they felt were the most appropriate. The results showed that all experts showed a common preference for situationally appropriate strategies. Moreover, these experts suggested that the criteria for identifying situationally appropriate strategies should include factors such as ease of calculation, fewer steps, and the goal of taking advantages of a problems’ characteristics. The result indicated that mathematics experts’ strategy evaluation criteria were consistent with previous studies that emphasized adaptive expertise, meta-strategic knowledge, as well as those on flexibility (Baroody, 2003; Star & Rittle-Johnson, 2008; Wang et al., 2019; Zohar, 2012).

Evaluating the Appropriateness of Strategies: A Subjective Perspective

Despite the centrality of the task context in the explicit evaluation of strategy appropriateness, another stream of research has focused on more implicit and subjective variables that are related to strategy use and strategy evaluation. According to strategy choice models (Marewski & Schooler, 2011; Shrager & Siegler, 1998; Siegler & Araya, 2005; Siegler & Shipley, 1995), individuals store information about a strategy’s cost (e.g. the time involved in adopting the strategy) and benefit (e.g. its accuracy in solving a problem) that is drawn from their problem-solving experiences. This information helps learners to continually adapt their strategy choices and eventually to find the best cost–benefit trade-offs—or in our context, identifying and then using the most appropriate strategies.

During such an adaptive learning process, if a strategy is task-relevant and if it is judged to be accurate and useful for the task, then this strategy is used. If a strategy is inaccessible or task-irrelevant, or if it is judged to be inaccurate, then a backup strategy is used to solve the problem. Over time, repeated exposure to these types of learning experiences will improve individuals’ strategy efficiency (Shrager & Siegler, 1998). However, strategy choice models argue that this process of strategy use and evaluation tends to be implicit and unconscious, as compared to the more explicit meta-strategic perspectives. These models suggest that an appropriate or adaptive strategy tends to have the strongest association with a problem and yields the greatest confidence on its utility and accuracy. Since both associative strength and confidence are mainly based on one’s own experience, the criteria of appropriateness can vary considerably among different leaners (Geary et al., 1993; Imbo & LeFevre, 2009).

This idea—that there may be substantial individual variation in learners' identification of the most appropriate or adaptive strategy—is further supported by studies that used a choice/no-choice paradigm (Siegler & Lemaire, 1997). Such studies have found that when participants were allowed to choose different strategies to solve a computational estimation task, their speed and accuracy were better than when they were asked to use a particular strategy to solve all the problems (Lemaire & Lecacheur, 2002; Siegler & Lemaire, 1997). The result implies that people can select the most accurate and fast strategy to solve a problem based on their experience. Similar findings have been also reported in other mathematics domains such as simple addition and subtraction (Fagginger Auer et al., 2016; Torbeyns & Verschaffel, 2013; Torbeyns et al., 2004), numerosity judgment (Luwel et al., 2005), and cognitive strategy use (e.g. retrieval and counting strategies) in arithmetic tasks (Imbo & Vandierendonck, 2007; Imbo et al., 2007). Taken together, these results indicate that people are capable of (unconsciously) choosing appropriate strategies for solving different problems if they are given options. However, such appropriateness is mainly based on one's experience and does not necessarily align with task goals or take advantage of the task characters.

Strategy selection models help to explain the subjective and diverse evaluation criteria used by learners in determining the appropriateness of strategies in equation solving (Franke & Carey, 1997; Star & Madnani, 2004). A previous study found that students' criteria for determining the best strategy in equation solving were quite complex. Although many students did express the belief that the strategy that involved the fewest steps was better (situationally appropriate), they also considered problem-solving accuracy and their own confidence in using the strategy when evaluating the merit of a strategy (Star & Madnani, 2004). Furthermore, students have also been found to exhibit idiosyncratic criteria when asked to select a situationally appropriate strategy for a given problem (Xu et al., 2017). They found that students usually can provide multiple strategies (which often included the situationally appropriate ones) to solve an equation, but some students did not indicate a preference for a situationally appropriate strategy (Star & Rittle-Johnson, 2008; Star et al., 2015a, b; Wang et al., 2019; Xu et al., 2017).

One possible explanation for students' lack of flexibility might be related to students' prior familiarity with the different strategies used in equation solving. Recent studies have found that students have a strong tendency to use the standard strategy in their initial problem-solving attempts, and this tendency might reflect students' familiarity with, and their confidence in, using known and accurate strategies (Liu et al., 2018; Wang et al., 2019; Xu et al., 2017). According to strategy selection models, it might also be the case that, when given a choice of strategies, students believe that whichever strategy they initially thought of and chose to use first is more appropriate, even if this strategy is not situationally appropriate.

Furthermore, such a consideration has the added effect of both affirming an individual's strategy choice (e.g. "the strategy is the best because it is the one that I thought of first") and self-reinforcing (e.g. thinking of a strategy first increases the likelihood that it will be considered first on future problems). Such self-referencing thoughts are usually involved when people evaluate different options (Brown et al., 1986; Dahl & Hoeffler, 2004; Sood & Forehand, 2005; West et al., 2004). In particular, the self-referential process refers to the evaluation of a choice or a decision in conjunction with or in relation to the individual's prior thoughts and experiences. Studies have suggested that people tend to show a more positive attitude toward a choice or a decision that is related to what they have experienced in the past as compared to the one that is unrelated to their experience (Brown et al., 1986; Sood & Forehand, 2005). Thus, in the case of equation solving, students might consider their first strategy as the best because this choice reflects their preference in strategies and their prior adaptive learning experience.

This subjective perspective on strategy appropriateness provides some insight into individual differences in strategy practice and evaluation. However, the self-referencing that underlies this subjective process may not align with task goals related to efficient or situationally appropriate strategies. The resulting strategy choices might lead to stereotyped and inflexible perseverance, including perseveration in using strategies that were successful in prior tasks but are no longer efficient or appropriate in later tasks (Carr & Steele, 2009). Such inflexibility might hinder the pursuit of more flexible knowledge and the development of adaptive expertise that is emphasized by educational researchers. Therefore, it is important to examine the extent to which students use subjective criteria when evaluating strategies. Furthermore, teachers' instruction (Jitendra et al., 2007; O'Sullivan & Pressley, 1984; Sawyer et al., 1992) and textbook design (Foxman, 1999; Horsley & Sikorová, 2015; Tian et al., 2021) have a major impact on students' strategy practice and experience that might affect the self-referencing process. As a result, it is also important to understand how different educational settings might affect students' strategy use and evaluation.

Cross-National Comparisons on Students' Problem-Solving Strategy

Cross-national studies provide researchers with special opportunities to understand how different cultures and educational settings affect students' mathematical achievement. Based on the results of cross-national research studies, we can design intervention programs to improve students' learning (Cai, 2004; Robitaille & Travers, 1992). Prior studies have found that students' use of strategies varies considerably between countries (Cai, 2004; Jiang et al., 2014, 2017; Xu et al., 2014; Yeap et al., 2006). In the field of algebra learning, recent international studies also found that students from different countries showed some similarities but also great diversity in strategy use when solving algebraic problems (Kilhamn & Säljö, 2019; Oikarainen, 2013; Reinhardtsen & Givvin, 2019; Röj-Lindberg & Partanen, 2019). For example, Reinhardtsen and Givvin (2019) investigated the problem-solving solutions to a match-stick patterning task among students from Finland, Norway, Sweden, and the USA. They found that students from the four countries used a similar drawing strategy to solve the task; however, only Swedish students used algebraic equations to solve the problem. Another study used different types of algebra problems and found that in the tasks where students were supposed to write arithmetic or algebraic expressions or equations, Finnish students generally used the typical mathematical procedures while Norwegian students present more atypical, but correct, solutions (Oikarainen, 2013). These studies suggest that cross-national comparison studies might be fruitful for illustrating variation in students' strategy usage, particularly for algebraic tasks. However, many of these studies only included a limited number of students for qualitative analysis and did not provide a detailed analysis of students' strategy evaluation or strategy flexibility.

With regard to strategic flexibility in equation-solving, previous studies have demonstrated that learning environments (such as classroom instruction or curriculum use) can have a considerable impact on students' flexible use of strategy (Rittle-Johnson & Star, 2007; Yakes & Star, 2011). Taking a broader perspective, teaching methods, educational systems, or even culture may also influence students' strategic flexibility since all these factors shape the learning environment. Studies that focused on the international comparison of algebra educational settings showed that the algebra curriculum and textbooks from different countries (including China, Finland, Norway, Spain, Sweden, and the USA) differ with regard to mathematics content as well as expected mathematical reasoning and cognitive levels (An et al., 2012; Givvin et al., 2019; Vicente et al., 2020). Such differences could further influence students' strategy evaluation and their flexible mathematical knowledge. However, rarely has published research investigated strategic flexibility while also considering the educational and cultural context of the different countries.

According to Verschaffel et al. (2009), sociocultural factors such as educational goals and classroom learning opportunities have powerful effects on students' strategy use and strategy knowledge. Researchers also have emphasized that we should pay more attention to the influence of students' educational histories and current instructional settings on their strategy choices (Baroody & Dowker, 2013; Bisanz, 2003). To further understand how different educational settings might affect students' mathematical thinking, the present study aimed at investigating students' strategy evaluation in Sweden, Finland, and Spain.

The Present Study

The present study focused on investigating students' strategy evaluation in equation solving from a cross-national perspective. We first examine the extent to which students use subjective evaluation criteria to identify the best strategy. If students can use the most efficient and accurate strategies based on an implicit association process (Shrager & Siegler, 1998) and as a result of a self-referencing process when determining the most appropriate strategy choice (Sood & Forehand, 2005), they should show a greater tendency to evaluate the strategy they use first as the most appropriate strategy. Yet at the same time, different educational settings might affect both students' strategy practices and their evaluation of different strategies in equation solving. For example, previous studies showed that the algebra curriculum and textbooks in Sweden and Finland tend to be diverse and inquiry-based while Spanish ones were focused more on typical algorithms (Kilhamn & Säljö, 2019; Vicente et al., 2020; Yang & Sianturi, 2020). Thus, we also examined whether a more varied and flexible learning environment (e.g. in Sweden) would encourage students to show a more flexible strategy evaluation pattern as compared to an algorithm-oriented environment would do (e.g. in Spain). Finally, the present study included middle school and high school students to allow for the examination of the impact of prior knowledge on students' flexibility (Schneider et al., 2011; Star & Seifert, 2006). Since high school students tend to have more learning experience and more practice on algebraic problem solving than middle school students, our third purpose is to examine whether such additional learning experience would be associated with increases in flexible knowledge of strategy evaluation.

Method

Participants

There were 792 middle school and high school students who participated in this study, including 288 Swedish students (87 in 9th grade, 201 in 11th grade), 258 Finnish students (93 in 8th grade, 165 in 11th grade), and 246 Spanish students (164 in 8th grade, 82 in 11th grade). We refer to 8th and 9th grade students as being in middle school (MS), and 11th grade as being in high school (HS). Participating students attended seven Swedish schools, eight Finnish schools, and five Spanish schools. The schools varied in both school size and geographic location. High school students were selected from both advanced (HS-A) and basic mathematics (HS-B) tracks; the proportion of the advanced track students in the study as compared to all participating high school students in the same country was 90% in Sweden, 63% in Finland, 57% in Spain. All the participants provided informed consent (or guardian's informed consent) and participated in accordance with national and international norms governing the use of human research participants.

Measures

The assessment used in this study was very similar to a flexibility assessment from previous studies (Liu et al., 2018; Wang et al., 2019; Xu et al., 2017). Our assessment was designed originally in English and Chinese, and it was subsequently translated into Swedish, Finnish, and Spanish by research collaborators who were native speakers in each country. The assessment included twelve equations (see Table 2), and each equation was designed to be solvable by a standard strategy and a situationally appropriate strategy. For each equation, we provided empty boxes for students to write several (up to six) strategies.

Data Collection Procedures

The same data collection procedures were applied in all countries.[1] Participants completed the test individually during a regular math class (45 min). To assess

[1] The dataset that we used in this study was published in Jiang, R., Star, J. R., & Hästö, P. (2022, May 16). Exploring students' procedural flexibility in three countries. https://doi.org/10.17605/OSF.IO/KWQDB.

Table 2 Sample items on the equation test

Sample equation	Standard strategy	Situationally appropriate strategy
$4(x-2)=24$	$4x-4\times2=24$ $4x-8=24$ $4x=32$ $x=8$	$x-2=24\div4$ $x-2=6$ $x=8$
$4(x+6)+3(x+6)=21$	$4x+24+3x+18=21$ $7x+42=21$ $7x=-21$ $x=-3$	$7(x+6)=21$ $x+6=21\div7$ $x=3-6$ $x=-3$
$8(x-5)=3(x-5)+20$	$8x-40=3x-15+20$ $8x-3x=5+40$ $5x=45$ $x=9$	$8(x-5)-3(x-5)=20$ $5(x-5)=20$ $x-5=4$ $x=9$
$\frac{5x+5}{5}+\frac{6x+6}{6}=6$	$30\times\left(\frac{5x+5}{5}\right)+30\times\left(\frac{6x+6}{6}\right)=30\times6$ $6\times(5x+5)+5\times(6x+6)=180$ $30x+30+30x+30=180$ $60x=120$ $x=2$	$\frac{5(x+1)}{5}+\frac{6(x+1)}{6}=6$ $(x+1)+(x+1)=6$ $2x+2=4$ $x=2$

students' strategy use and strategy evaluation, we applied a Tri-Phase procedure that was found to be valid and reliable in measuring students' strategy flexibility (Xu et al., 2017). In Phase 1, students were asked to solve the twelve equations as quickly and accurately as possible, and to write down their solution on the first box in the answer sheet. This phase aimed at probing students' first strategy attempt. We provided 15 min for students to complete this phase. In Phase 2, students were asked to solve each equation by using as many different strategies as possible that were written in the provided boxes. Students were provided 20 min to complete this phase. This phase aimed at encouraging students to generate multiple strategies. In Phase 3, students were asked to review their strategies and select the strategy that they felt was the best one for each equation (by circling the box for that strategy). Students were given 5 min for this phase. This phase aimed at probing students' strategy evaluation.[2]

Coding

Students' strategies were coded into four strategy types. With respect to the reliability of this strategy coding, for the Swedish sample, 5% of the student cases (the number of equations on the test (12) × number of students) were coded by two native Swedish speakers; the intercoder agreement was 98.5%. For the Finnish sample, 5% of the student cases were coded by two native Finnish speakers, with intercoder agreement at 98.3%. For the Spanish sample, all student cases were coded by two

[2] The dataset that we used in this study was published on https://doi.org/10.17605/OSF.IO/KWQDB.

native Spanish speakers, with intercoder agreement of 98.6%. All disagreements between the two coders were resolved by discussion.

Students' strategies for each equation were coded into four categories: standard, situationally appropriate, other, and no/unintelligible strategy. The standard and situationally appropriate strategies are presented in Table 2. Strategies coded as other were neither standard nor situationally appropriate but were potentially viable strategies that could be used to solve the equation correctly (see Table 7, the third box of the iC + uA example). Strategies that involved unintelligible work (e.g. only rewriting the equation) and empty boxes were coded as no strategy. Note that accuracy was not considered as part of the strategy coding process.

In addition, the order of students' strategy use was determined based on which strategy was written in which box for each problem. For example, if a strategy was written in the first box, then the strategy was coded as the "first box strategy." Finally, students' strategy selections in Phase 3 were coded based on the number of the box that was circled for each problem. For example, if a student circled the strategy that was written in the first box for a problem, then this first box strategy was also identified as the circled box strategy. The coding of the order in which strategies were used and the strategy that was circled by the student for each problem were used to determine which strategy students believed to be appropriate, as detailed below.

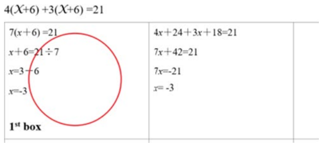
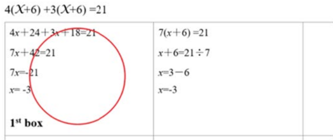
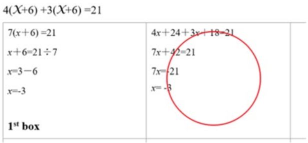
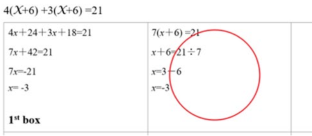
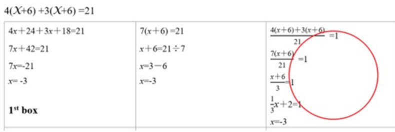
Analysis

Given our interest in understanding students' strategy choices and rationales for evaluating strategies, our analysis began by identifying our first analytical sample, which included all cases where (a) a student produced a solution in Phase 1 of the assessment and at least one solution in Phase 2, and also (b) the student indicated (circled) which solution was best in Phase 3. To enable our evaluation of students' strategy choices, it was necessary to analyze only those cases where students produced multiple strategies and also selected the one strategy that they perceived to be the best.

Within this first analytical sample of problems, we defined several strategy evaluation types by determining whether a student's strategy choices for a problem were consistent (C) or inconsistent (iC). The concept of "consistency" here is borrowed from the literature of dual-process theory that focuses on heuristic and analytical cognitive processes.[3] As suggested by strategy choice models, people tend to select their first strategy for solving a problem based on their strategy practice or experience (Shrager & Siegler, 1998). Thus, if students justify their first strategy as the

[3] Heuristic processes are based on experience/intuition and are automatic and more implicit (i.e. the rule of thumb), while analytical processes are based on logic and deliberate thinking. People's responses or tasks that are aligned with experience are usually considered as consistent responses or consistent tasks, and responses or tasks that require deliberate reasoning to overcome the heuristic are considered as inconsistent responses or inconsistent tasks (Chaiken & Ledgerwood, 2012; De Neys, 2006).

Table 3 Different strategy evaluation types and examples of the students’ responses

	A (Aligned with flexibility goals)	uA (Not aligned with flexibility goals)
C (consistent)	C+A: 	C+uA: 
iC (Inconsistent)	iC+iA:  iC+eA: 	iC+uA: 

best strategy, their responses were considered as consistent. Operationally, if a student circled (in Phase 3) the strategy written in the first box that was generated in Phase 1, the student was coded as consistent (C) for that problem. In other words, C indicated the “self-referencing” process as noted above. Otherwise, if a student circled any other strategy for a problem apart from the one written in the first box, the student was coded as inconsistent (iC). The C and iC responses reflect students’ prior adaptive learning experience as well as their initial preference and confidence in strategy use.

Our second aim was to investigate the alignment between students’ strategy choices and that emphasized in the flexibility literature. For this aim, it was necessary to use a subset of the first analytic sample, only including cases where (a) the strategies that a student produced for a given problem (in Phases 1 and 2) included both a standard and a situationally appropriate solution, and (b) the first strategy generated (for Phase 1) was one of these two strategy types (i.e. not of type “other”).

Within this second analytic sample, external alignment captured the extent to which a student’s strategy choices aligned with the flexibility views. Students expressed their preferences for which strategies were best through their choice of which strategy to use in the first box in Phase 1 (we view this as an implicit identification of a best strategy) as well as which strategy to circle as the best in Phase 3 (we view this as an explicit identification of a best strategy). Thus, for a given problem, a student’s strategies could be aligned with flexibility goals implicitly (the implicit process suggested by the strategy choice models), where the student chose to use a strategy as their first attempt in Phase 1 that is defined as the best strategy in the flexibility literature (iA), or explicitly, where the student circled a strategy in Phase 3 that flexibility researchers viewed as the best strategy (eA). Alternatively, students' indication of which strategy is best for a problem (implicitly or explicitly) might not align with the goal of flexibility (uA).

These two types of coding—which indicate (a) consistency between students' first strategy and the one that was circled as the best strategy, and (b) alignment between students' identification of best strategy and the goals that flexibility literature emphasized—can be combined into the matrix described as follows (see Table 3). Note that it was theoretically possible for a student to be non-internally consistent and not externally aligned (iC + uA). But in practice, such cases were extremely rare (8 out of 1945, or less than 0.5%) and as a result were excluded from our analyses. All the coded data were analyzed by chi-square analysis, and level-wise comparisons were performed to test differences among age groups or country groups. Bonferroni correction was used when multiple tests were performed on the same variables.

Results

In this section, we describe our first analytical sample, which included all cases where a student provided strategies in both Phases 1 and 2 and circled a choice in Phase 3. Within this analytical sample, we then analyze the frequency of C and iC cases to determine to what extent students' responses were in accordance with the predictions of the strategy choice models. We analyze the four types of strategy evaluation responses (see Table 3) in the second analytical sample, namely C + A, iC + iA, iC + eA, and C + uA, to examine how students' evaluation of strategies might relate to their strategy preferences and to what extent their criteria for strategy evaluation aligned with the goal of flexibility development.

Internal Consistency

We now restrict our attention to the first analytical sample (i.e. those cases that were deemed either C or iC). The overview of the data and the construction of the first analytical sample were presented in Table 4.

Overall levels of internal consistency (C) and inconsistency (iC) responses among our first analytical sample from each group of students are presented in Table 5. Note that there were some cases with multiple solutions who were neither C nor iC. In such cases, students have either presented no solution in Phase 1 or they have not indicated which solution they consider the best.

The number of cases in the analytical sample ranges from 14.0 (MS, Spain) to 52.6% (HS-A, Sweden) as shown in Table 4. However, Table 5 indicates that in each group the relative frequency of C and iC cases are roughly the same, ranging from 48.0 (HS-A, Finland) to 66.7% (MS, Spain) with the other groups in the range 50–60%. Thus, students from all countries showed a similar percentage of responses of circling the first box strategy as the best. With regard to age differences pooled

Table 4 Rates of multiple strategy use and consistency among all cases in for Spanish, Finnish, and Swedish students in different tracks

Country	Age group	Percent of cases where multiple strategies were used	Percent of cases that were consistent (C)	Percent of cases that were inconsistent (iC)	Percent of cases included in the first analytical sample
Spain	MS ($n=1968$)	16.0	9.3	4.7	14.0
	HS-B ($n=420$)	38.1	22.6	15.2	37.9
	HS-A ($n=564$)	48.2	26.2	20.4	46.6
Finland	MS ($n=1116$)	14.3	6.7	6.5	13.3
	HS-B ($n=732$)	19.1	10.0	7.0	16.9
	HS-A ($n=1236$)	53.2	25.0	27.1	52.1
Sweden	MS ($n=1044$)	24.2	10.5	9.8	20.3
	HS-B ($n=240$)	40.0	20.4	19.2	39.6
	HS-A ($n=2172$)	53.8	29.6	23.0	52.6

MS middle school, *HS-B* high school basic track, *HS-A* high school advanced track.

n is the total number of cases for each group, which is the number of equations on the test (12) × number of students.

Table 5 The internal consistency rate (%) in the first analytical sample for Spanish, Finnish, and Swedish students in different tracks

Country	Age group	Percent of cases that were internally consistent (C)
Spain	MS ($n = 276$)	66.7
	HS-B ($n = 159$)	59.7
	HS-A ($n = 263$)	56.3
Finland	MS ($n = 148$)	50.7
	HS-B ($n = 124$)	58.9
	HS-A ($n = 644$)	48.0
Sweden	MS ($n = 212$)	51.9
	HS-B ($n = 95$)	51.6
	HS-A ($n = 1143$)	56.3

n is the total number of cases that students proved either a consistent or inconsistent response.

over the three countries, there was not a statistically significant difference in rates of internal consistency χ^2 (2, 3046) = 4.7, $p = 0.095$. In total, there were 1686 cases of internal consistency, for a rate of 55.5% of cases.

We move now to a comparison of the results from the three different countries, in terms of rates of internal consistency (i.e. percent of cases where a student circled their first strategy as the best). Since the composition of the samples from different countries differs, we compare the countries within track (MS, HS-B, or HS-A). We find that there is no difference for HS-B, χ^2 (2, 378) = 1.8, $p = 0.41$, whereas the other two tracks do feature significant between-country differences (MS: χ^2 (2, 636) = 15, $p = 0.001$, and HS-A: χ^2 (2, 2050) = 12, $p = 0.002$). Follow-up *t* tests with Bonferroni correction show that middle school Spanish students have higher rates of C cases compared to Finland and Sweden; the latter two are not different from one another. For HS-A, Finnish students have lower rates of C cases compared to their Spanish and Swedish peers, who do not differ from one another.

In sum, with regard to the internal consistency of students' identification of which strategy was the best, our results suggest that it was not the case that students tended to use the strategy that they viewed as the best in their first attempt. In only about 55.5% of all cases students were internally consistent. Furthermore, this proportion of cases for which students were internally consistent did not change significantly from middle school to high school. Spanish middle school students were slightly more internally consistent than their peers in the other countries, whereas Finnish advanced track high school students were slightly less internally consistent than their peers. These results indicate that students' preferred strategy—the one that they chose to use first on each problem—was not necessarily the one that they viewed as being the best.

Table 6 Different types of strategy evaluation (%) among all the cases for Swedish, Finnish, and Spanish students in different tracks

Country	Age group	C + A	iC + iA	iC + eA	C + uA	Proportion of all cases included in the 2nd analytical sample
Spain	MS (*n* = 1968)	0.2	0.1	2.1	2.5	4.9
	HS-B (*n* = 420)	2.6	0.5	11.7	5.5	20.5
	HS-A (*n* = 564)	1.4	0.0	16.0	8.2	25.5
Finland	MS (*n* = 1116)	1.3	0.1	3.0	0.5	5.0
	HS-B (*n* = 732)	0.1	0.0	2.9	2.2	5.3
	HS-A (*n* = 1236)	12.3	0.8	23.1	6.1	42.3
Sweden	MS (*n* = 1044)	0.4	0.2	5.4	2.5	8.4
	HS-B (*n* = 240)	2.1	0.4	14.2	5.4	22.1
	HS-A (*n* = 2172)	18.2	1.2	17.1	2.9	39.5

Table 7 Different types of strategy evaluation (%) among the second analytical sample for Swedish, Finnish, and Spanish students in different tracks

Country	Age group	C-A	iC-iA	iC-eA	C-uA
Spain	MS (*n* = 97)	4.1	1.0	42.3	51.5
	HS-A (*n* = 144)	5.6	0.0	62.5	31.9
Finland	MS (*n* = 56)	26.8	1.8	60.7	10.7
	HS-A (*n* = 523)	29.1	1.9%	54.5	14.5
Sweden	MS (*n* = 88)	4.5	2.3	63.6	29.5
	HS-A (*n* = 859)	46.1	2.9	43.2	7.2

n is the total number of cases from the second analytical sample that students proved both a standard and a situationally appropriate solution and the first solution is of one of these types.

Strategy Evaluation

To further examine students' evaluation of strategies and to explore the extent to which their criteria for strategy evaluation aligned with flexibility goals, we analyzed the four types of strategy evaluation responses, namely C + A, iC + iA, iC + eA, and C + uA. The percentages of each type of strategy evaluation from each group of students are presented in Table 6.

From the final column, we see that middle school and Finnish high school basic track students had only 4.9–8.4% of cases included in the analytical sample, high school students in either track in Spain and the basic track in Sweden were in the range 20.5–25.5%, and advanced track high school students in Finland and Sweden included 42.3% and 39.5% of cases. This pattern is similar to that of the proportion of cases including multiple strategies (see Table 4), with the exception of Spanish advanced track high school students who are relatively less represented in the second sample than the above analysis.

Furthermore, we observe that the iC + iA case is extremely rare, with less than 1% of cases in all countries and age groups. The C + A case is also quite rare, with the exception of advanced high school track students in Finland. In addition, high school basic track is not well represented in the second analytical sample for Finnish students, with only 5.3% of the cases included in the sample. Although other countries had a modest rate of HS-B included in the second analytical sample (20.5% and 22.1% for Spain and Sweden, respectively), the original sample is relatively small. In these circumstances, a detailed analysis of the categories can easily be swayed by individual students, so we omit the HS-B group in what follows (Table 7).

Across the three countries, there was a significant difference between middle school and high school advanced track students, χ^2 (3, 1822) = 118, $p < 0.001$. A follow-up comparison of groups within country (with Bonferroni correction coefficient 3) showed significant differences in Spain, χ^2 (3, 243) = 11.3, $p < 0.032$, no significant differences in Finland, χ^2 (3, 591) = 1.43, $p > 0.1$, and significant differences in Sweden, χ^2 (3, 988) = 118, $p < 0.001$. Pair-wise comparisons of groups across countries with a Bonferroni coefficient 15 (equal to the number of pairs among the six groups) suggested that Spanish HS-A and Swedish MS are very similar. A comparison of either of these with either the Finnish group shows significant differences, e.g. Finnish and Swedish MS comparison gives χ^2 (3, 151) = 21.6, $p < 0.002$ (with correction). The difference between Swedish HS-A and the others is even bigger ($ps < 0.001$).

In summary, Swedish advanced track high school students showed the strongest tendency to identify the situationally appropriate solution as best: 90% of cases were aligned with the flexibility goals. In half of these cases, the situationally appropriate solution was produced in Phase 1 (C + A); in the other half, it was produced in Phase 2 (iC + eA). Most of the remaining cases are of type C + uA, which typically involved using the standard algorithm in Phase 1 and selecting it as the best. Finnish students in middle school or the advanced track of high school also accounted for 84–88% of the cases that selected the situationally appropriate solution as best. However, as compared to their Swedish high school peers, only one-third of cases involving situationally appropriate solutions were produced in Phase 1 (C + A), whereas production of the situationally appropriate solution in Phase 2 (iC + eA) accounted for two-thirds of cases. Again, most of the remaining cases are of type C + uA.

As compared to Swedish high school students and Finnish students, Spanish and Swedish middle school students showed a very different pattern when evaluating the best strategy. For these students, the type C + uA was now present in 30–52% of the cases. This indicates Spanish and Swedish middle school students were much less likely (as compared to the other profiles) to have chosen the situationally appropriate solution as the best when it was not produced in Phase 1. Indeed, for Spanish middle school students, more than half of all cases involved starting with the standard solution and choosing it as best, even though these students were able to produce the situationally appropriate solution in Phase 2. Overall, in only 46–68% of cases was the situationally appropriate solution chosen as best for this group of students.

Our results also indicated that each country showed a different pattern of change from middle school to high school. Finnish middle school students and advanced track high school students showed no difference in the distribution of cases. In Spain, there were indications of a shift ($p < 0.032$), with more high school students being internally inconsistent in favor of choosing the situationally appropriate solution produced in Phase 2. Finally, in Sweden there was a substantial difference between the age groups, with a shift from internally consistent and non-aligned with flexibility goals (C + uA) to internally consistent and aligned with flexibility goals (C + A).

Discussion

The present study explored students' strategy evaluation in three different countries: Sweden, Finland, and Spain. To our knowledge, this is the first cross-national study that examined the issue of strategy evaluation in equation solving. Through the development of a new coding system of students' responses on flexibility tasks, we accessed both students' implicit strategy choices for how to solve equations as well as their explicit evaluation of the appropriateness of strategies. Specifically, we examined how students' evaluation of which strategies were considered the most appropriate might relate to their adaptive learning experience and strategy selection preferences as suggested by strategy choice models and the self-referencing account of choice making, to what extent students' strategy evaluation choices were aligned with the flexibility literature, and how students' strategy evaluation choices might differ based on different national contexts and age levels.

Our results indicated that, first, in cases where students used multiple strategies to solve linear equations on our assessment, there was no clear indication that students viewed their first strategy as the best strategy for a given problem. In about half of the cases, students identified a strategy other than the one that they used first as the best strategy (i.e. iC). This pattern did not differ significantly across countries or between middle school and high school. This result is consistent with previous studies' findings which affirm that students do not always use the most appropriate strategies even if they demonstrated knowledge of them (Liu et al., 2018; Xu et al., 2017). Furthermore, these results challenge the strategy selection models' claim that students can select the most adaptive/efficient strategy to solve a problem as long as they are given choices. That is, the strategy that students use in their first attempt might indeed reflect their familiarity or confidence in using that strategy (as predicted by strategy choice models). But it is difficult to claim that students' first used strategy is the one that they considered to be the best or the most appropriate, given that many students often did not evaluate their first-used strategy as the best. In addition, in high percentages of cases in both middle school and high school, students across all three countries indicated that they did not appear to be self-referencing by selecting their first strategy as the best regardless of what type of strategy it was. Our findings suggest that students from all countries evaluated strategies not only based on their initial preferences or their confidence of using a strategy, but also using other factors such as the innovativeness or efficiency of the strategy, or perhaps their beliefs about the teacher's expectations. Such findings support the idea that students could show explicit meta-strategic knowledge when evaluating the appropriateness of strategies (especially for higher-track students) instead of relying on the implicit and automatic associative process (Luwel et al., 2003).

Second, when considering the alignment between students' and flexibility goals on the evaluation of strategies, it appeared that students' strategy evaluation patterns were quite varied across countries. We found that Finnish and Swedish students' strategy evaluation criteria were more likely to be aligned with researchers' criteria, as compared to Spanish students. This was even more pronounced for Swedish advanced track high school students. This strategy evaluation pattern is also consistent with prior studies that included only Chinese students (Wang et al., 2019; Xu et al., 2017). Spanish students and Swedish middle school students, however, tended to use the standard strategy in their first attempt and to evaluate the standard strategy as the best. We believe that these cross-national differences in strategy use and strategy evaluation might be related to the different educational environments in three countries, particularly in the ways that recent changes in the Finnish and Swedish mathematics curricula may have moved these countries away from the curricular model used in Spain. In particular, Spanish researchers have pointed out that many existing Spanish textbooks continue to include many traditional activities (e.g. pen-and-pencil exercises and many algorithmic exercises) and do not leverage various contexts or methods that could help students understand mathematics concepts (Rico, 1995; Socas et al., 2016; Vega-Castro et al., 2012). Commentators on Spanish mathematics teaching and curriculum note that linear equation solving consists mostly of routine exercises and mathematical tasks are almost always closed and with a unique solution, and teachers' task implementation decisions do not promote students use of multiple strategies and the explicit comparison of different solutions to the same problem (e.g. González-Astudillo & Sierra-Vázquez, 2004; López & Betancor, 2007). Furthermore, a majority of teachers still rely heavily on quite traditional textbooks and continue to place great emphasis on using standard algorithms quickly and efficiently (Casas & Garcia Castellar, 2004; Joglar-Prieto et al., 2018). In particular, the present results suggest that Spanish students might value standard strategies more than the situationally appropriate strategies, to the point where they tend to evaluate the standard strategy as the best.

While the traditional approach from Spain is likely also present in Finland and Sweden, it is also the case that there also have been substantial efforts to encourage teachers to use open-ended mathematics problems in Finland during the past decades (Halinen et al., 1991; Pehkonen, 2008). A similar trend of an increasing emphasis on understanding mathematics, interpreting mathematical situations, and using strategies to solve problems has also been observed in Sweden (Bråting, 2015). In particular, there is more content that involves meaning-making, combined with a decrease in "pure" mathematics content (e.g. calculation practice) (Bråting, 2015; Hemmi et al., 2013). These efforts in Sweden and Finland have possibly caused some increase in meaningful contexts and open-ended questions. If students are given freedom in the problem-solving process to understand and explore different strategies, students may end up producing multiple and equally correct solutions, depending on the problem-solving choices made and the emphases placed during their solution processes (Pehkonen, 2008). This kind of educational environment may encourage students to use different strategies and think creatively during problem-solving, which in turn may further impact students' beliefs about what it means for a strategy to be the best strategy. These cross-national differences in strategy evaluation further support the idea that students evaluate the appropriateness of a strategy not only based on task characteristics (e.g. the structural features of an equation) or their practices (e.g. how familiar and confident they are in using the strategy), but also on variables related to sociocultural context and classroom teaching and learning norms (Ellis, 1997; Verschaffel et al., 2009).

Finally, we found that different countries showed different strategy evaluation patterns regarding age differences (i.e. middle school and high school). In Spain and Sweden, high school students were more likely to identify the situationally appropriate strategy as the best as compared to middle school students, as evidenced by an increase in the number of C + A (Sweden) and iC + eA (Spain) cases from middle school to high school. In contrast, Finnish students were an exception to this trend with both middle and high school student groups lying somewhere in between Spanish and Swedish high school students. We conjecture that the anomalous result of the Finnish middle school group may be explained by the fact that the group included only 56 cases, so the results may have been swayed by just a few especially talented students.

In general, high school students have more algebraic knowledge than middle school students; they also have more equation-solving experience. According to previous studies, students with greater prior knowledge were more capable of solving equations in multiple ways and are more likely to recognize the strategies that fits the task context to be appropriate (Newton et al., 2019; Schneider et al., 2011; Verschaffel et al., 2011). Such age differences might also be related to other cognitive factors such as executive function, metacognition, or meta-strategic knowledge because these factors are age sensitive and play important roles in problem solving (Diamond, 2013; Zohar, 2012). It would be interesting to further investigate whether students' evaluation is related to these cognitive factors.

In sum, the present study suggests that how students evaluate a situationally appropriate strategy as the best strategy may be influenced by their sociocultural and country context and their prior knowledge and age level. Previous studies that dedicated to understanding, characterizing, and promoting students' mathematics flexibility emphasized the ability identifying the situationally appropriate strategy as the best strategy (Liu et al., 2018; Wang et al., 2019; Xu et al., 2017). However, when considering an educational context where the use of the standard strategy appears to be especially emphasized (e.g. in Spain), students might perceive that fast, accurate implementation of the standard algorithm continues to be a more important outcome than flexibility.

Furthermore, the cross-national differences in strategy evaluation found here also suggest that for students from different educational contexts, different types of interventions might be needed to improve students' flexibility. For example, for students who evaluate the situationally appropriate strategy as best but do not use it consistently or spontaneously, interventions that aim at improving students' understanding of equation structure and problem-solving procedures might be appropriate. However, for teaching environments where the standard strategy is emphasized, teacher professional development workshops that could help teachers understand the idea of procedural flexibility and the importance of using meta-strategic knowledge about the situationally appropriateness might be a more effective first step.

Although we found significant cross-national and developmental differences in strategy evaluation, there are several limitations to the present study that suggest caution in generalizing these conclusions. First, our study used convenience sampling instead of other systematic sampling methods. Thus, the distribution of our data across ages and countries is quite variable. Thus, our results might be affected by the sample size and quality. Future studies should use a more systematic sampling method to better control the sample size and other sample-related variables. Second, we used an age-match design in the present study and did not control for students' mathematics performance or ability. This might yield confounding effects on the cross-national comparison results. Prior studies have suggested that an ability-match design might be necessary because cross-national differences could be due to different levels of mathematics ability rather than cultural factors (Muldoon et al., 2011). Thus, future studies could use mathematic ability assessments or mathematics tasks to control for students' mathematics performance. Finally, this is a cross-sectional study which limits the generalization of our developmental claims. A longitudinal design might be necessary to better reveal the developmental trends of students' strategy evaluation.

Acknowledgements We thank all the teachers and students for their participation in this study.

Funding This study was also supported by a grant from the Guangdong Science and Technology Department (2021A1515010936).

References

An, T., Mintos, A., & Yigit, M. (2012). A cross-national standards analysis: Quadratic equations and functions. In L. R. Van Zoest, J.-J. Lo, & J. L. Kratky (Eds.), *Proceedings of the 34th annual meeting of the North American Chapter of the International Group for the Psychology of Mathematics Education* (Vol. 11, p. 120). ERIC (ED584829).

Australian Education Ministers. (2006). *Statements of learning for mathematics*. Curriculum Corporations.

Baroody, A. J. (2003). The development of adaptive expertise and flexibility: The integration of conceptual and procedural knowledge. In A. J. Baroody & A. Dowker (Eds.), *The development of arithmetic concepts and skills: Constructing adaptive expertise* (pp. 1–33). Lawrence Erlbaum Associates.

Baroody, A. J., & Dowker, A. (Eds.) (2013). *The development of arithmetic concepts and skills: Constructive adaptive expertise*. Routledge.

Bisanz, J. (2003). Arithmetical development: Commentary on Chapters 1 through 8 and reflections on directions. In A. J. Baroody & A. Dowker (Eds.), *The development of arithmetic concepts and skills: Constructive adaptive expertise* (p. 435). Lawrence Erlbaum Associates.

Bråting, K. (2015). Teaching traditions in Swedish school algebra–a project description. In I. Bjarnadottir, F. Furinghetti, J. Prytz, & G. Schubring (Eds.), *Proceedings of the Third International Conference on the History of Mathematics Education* (pp. 81–92). Uppsala Universitet.

Brown, P., Keenan, J. M., & Potts, G. R. (1986). The self-reference effect with imagery encoding. *Journal of Personality and Social Psychology, 51*(5), 897–906. https://doi.org/10.1037/0022-3514.51.5.897

Cai, J. (2004). Why do U.S. and Chinese students think differently in mathematical problem solving?: Impact of early algebra learning and teachers' beliefs. *The Journal of Mathematical Behavior, 23*(2), 135–167. https://doi.org/10.1016/j.jmathb.2004.03.004

Carr, P. B., & Steele, C. M. (2009). Stereotype threat and inflexible perseverance in problem solving. *Journal of Experimental Social Psychology, 45*(4), 853–859. https://doi.org/10.1016/j.jesp.2009.03.003

Carroll, W. M. (1999). Invented computational procedures of students in a standards-based curriculum. *The Journal of Mathematical Behavior, 18*(2), 111–121. https://doi.org/10.1016/S0732-3123(99)00024-3

Casas, A. M., & Garcia Castellar, R. (2004). Mathematics education and learning disabilities in Spain. *Journal of Learning Disabilities, 37*(1), 62–73. https://doi.org/10.1177/00222194040370010701

Chaiken, S., & Ledgerwood, A. (2012). A theory of heuristic and systematic information processing. In P. A. M. Van Lange, A. W. Kruglanski, & E. T. Higgins (Eds.), *A Handbook of Theories of Social Psychology* (pp. 246–266). SAGE.

Dahl, D. W., & Hoeffler, S. (2004). Visualizing the self: Exploring the potential benefits and drawbacks for new product evaluation. *Journal of Product Innovation Management, 21*(4), 259–267. https://doi.org/10.1111/j.0737-6782.2004.00077.x

De Smedt, B., Torbeyns, J., Stassens, N., Ghesquière, P., & Verschaffel, L. (2010). Frequency, efficiency and flexibility of indirect addition in two learning environments. *Learning and Instruction, 20*(3), 205–215. https://doi.org/10.1016/j.learninstruc.2009.02.020

DeCaro, M. S. (2016). Inducing mental set constrains procedural flexibility and conceptual understanding in mathematics. *Memory & Cognition, 44*(7), 1138–1148. https://doi.org/10.3758/s13421-016-0614-y

De Neys, W. (2006). Automatic–heuristic and executive–analytic processing during reasoning: Chronometric and dual-task considerations. *The Quarterly Journal of Experimental Psychology*, *59*(6), 1070–1100. https://doi.org/10.1080/02724980543000123

Diamond, A. (2013). Executive functions. *Annual Review of Psychology, 64*, 135–168. https://doi.org/10.1146/annurev-psych-113011-143750

Ellis, S. (1997). Strategy choice in sociocultural context. *Developmental Review, 17*(4), 490–524. https://doi.org/10.1006/drev.1997.0444

Fagginger Auer, M. F., Hickendorff, M., & van Putten, C. M. (2016). Solution strategies and adaptivity in multidigit division in a choice/no-choice experiment: Student and instructional factors. *Learning and Instruction, 41*, 52–59. https://doi.org/10.1016/j.learninstruc.2015.09.008

Foxman, D. (1999). *Mathematics textbooks across the world: Some evidence from the Third International Mathematics and Science Study (TIMSS)*. National Foundation for Educational Research.

Franke, M. L., & Carey, D. A. (1997). Young children's perceptions of mathematics in problem-solving environments. *Journal for Research in Mathematics Education*, *28*(1), 8–25. https://doi.org/10.5951/jresematheduc.28.1.0008

Geary, D. C., Frensch, P. A., & Wiley, J. G. (1993). Simple and complex mental subtraction: Strategy choice and speed-of-processing differences in younger and older adults. *Psychology and Aging, 8*(2), 242–256. https://doi.org/10.1037//0882-7974.8.2.242

Givvin, K. B., Geller, E. H., & Stigler, J. W. (2019). How teachers introduce algebra and how it might affect students' beliefs about what it means to "do" mathematics. In C. Kilhamn & R. Säljö (Eds.), *Encountering algebra* (pp. 139–163). Springer. https://doi.org/10.1007/978-3-030-17577-1

González-Astudillo, M. T., & Sierra-Vázquez, M. (2004). Textbook analysis methodology of mathematics: Critical points in secondary education in Spain during the twentieth century. *Science Teaching : A Review of Research and Teaching Experience, 22*(3), 705.

Halinen, H., Joki, L., Näätänen, P., Pehkonen, S., & Sainio, S. (1991). *On the trend of mathematics teaching in the comprehensive school in the 1990's*. VAPK-kustannus.

Hatisaru, V. (2021). A reflection on Star and Seifert's operationalisation of flexibility in equation solving. *For the Learning of Mathematics, 41*(2), 37–44.

Hemmi, K., Lepik, M., & Viholainen, A. (2013). Analysing proof-related competences in Estonian, Finnish and Swedish mathematics curricula—Towards a framework of developmental proof. *Journal of Curriculum Studies, 45*(2), 37–44. https://doi.org/10.1080/00220272.2012.754055

Horsley, M., & Sikorová, Z. (2015). Classroom teaching and learning resources: International comparisons from TIMSS – A preliminary review. *Orbis Scholae, 8*(2), 43–60. https://www.ceeol.com/search/article-detail?id=36707

Imbo, I., Duverne, S., & Lemaire, P. (2007). Working memory, strategy execution, and strategy selection in mental arithmetic. *The Quarterly Journal of Experimental Psychology, 60*(9), 1246–1264. https://doi.org/10.1080/17470210600943419

Imbo, I., & LeFevre, J.-A. (2009). Cultural differences in complex addition: Efficient Chinese versus adaptive Belgians and Canadians. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 35*(6), 1465–1476. https://doi.org/10.1037/a0017022

Imbo, I., & Vandierendonck, A. (2007). The development of strategy use in elementary school children: Working memory and individual differences. *Journal of Experimental Child Psychology, 96*(4), 284–309. https://doi.org/10.1016/j.jecp.2006.09.001

Impecoven-Lind, L. S., & Foegen, A. (2010). Teaching algebra to students with learning disabilities. *Intervention in School and Clinic, 46*(1), 31–37. https://doi.org/10.1177/1053451210369520

Jiang, C., Hwang, S., & Cai, J. (2014). Chinese and Singaporean sixth-grade students' strategies for solving problems about speed. *Educational Studies in Mathematics, 87*(1), 27–50. https://doi.org/10.1007/s10649-014-9559-x

Jiang, R., Li, X., Fernández, C., & Fu, X. (2017). Students' performance on missing-value word problems: A cross-national developmental study. *European Journal of Psychology of Education, 32*(4), 551–570. https://doi.org/10.1007/s10212-016-0322-9

Jiang, R., Liu, R.-D., Star, J., Zhen, R., Wang, J., Hong, W., Jiang, S., Sun, Y., & Fu, X. (2021). How mathematics anxiety affects students' inflexible perseverance in mathematics problem-solving: Examining the mediating role of cognitive reflection. *British Journal of Educational Psychology, 91*(1), 237–260. https://doi.org/10.1111/bjep.12364

Jitendra, A. K., Griffin, C. C., Haria, P., Leh, J., Adams, A., & Kaduvettoor, A. (2007). A comparison of single and multiple strategy instruction on third-grade students' mathematical problem solving. *Journal of Educational Psychology, 99*(1), 115–127. https://doi.org/10.1037/0022-0663.99.1.115

Joglar-Prieto, N., Abánades-Astudillo, M. Á., & Star, J. R. (2018). Flexibilidad matemática y resolución de ecuaciones lineales. *Uno: Revista de Didáctica de las Matematicas, 80*, 51–57.

Karlen, Y. (2015). Nutzungshäufigkeit von Lernstrategien und metakognitives Strategiewissen in der Oberstufe des Gymnasiums: Entwicklung und Zusammenhänge. *Zeitschrift für Bildungsforschung, 5*(2), 159–175. https://doi.org/10.1007/s35834-015-0123-2

Keleş, T., & Yazgan, Y. (2021). Gifted eighth, ninth, tenth and eleventh graders' strategic flexibility in non-routine problem solving. *The Journal of Educational Research, 114*(4), 332–345. https://doi.org/10.1080/00220671.2021.1937913

Kilhamn, C., & Säljö, R. (Eds.) (2019). *Encountering algebra: A comparative study of classrooms in Finland, Norway, Sweden, and the USA*. Springer. https://doi.org/10.1007/978-3-030-17577-1

Kuhn, D., & Pearsall, S. (1998). Relations between metastrategic knowledge and strategic performance. *Cognitive Development, 13*(2), 227–247. https://doi.org/10.1016/S0885-2014(98)90040-5

Lemaire, P., & Lecacheur, M. (2002). Children's strategies in computational estimation. *Journal of Experimental Child Psychology, 82*(4), 281–304. https://doi.org/10.1016/s0022-0965(02)00107-8

Liu, R.-D., Wang, J., Star, J. R., Zhen, R., Jiang, R.-H., & Fu, X.-C. (2018). Turning potential flexibility into flexible performance: Moderating effect of self-efficacy and use of flexible cognition. *Frontiers in Psychology, 9*, 646. https://doi.org/10.3389/fpsyg.2018.00646

López, N. R., & Betancor, J. B. (2007). Mathematics education in Spain [La educación matemática en España]. *Educational Praxis (Brazil)[Práxis Educativa (Brasil)], 2*(2), 151–160.

Luwel, K., Lemaire, P., & Verschaffel, L. (2005). Children's strategies in numerosity judgment. *Cognitive Development, 20*(3), 448–471. https://doi.org/10.1016/j.cogdev.2005.05.007

Luwel, K., Torbeyns, J., & Verschaffel, L. (2003). The relation between metastrategic knowledge, strategy use and task performance: Findings and reflections from a numerosity judgement task. *European Journal of Psychology of Education, 18*(4), 425–447. https://doi.org/10.1007/BF03173245

Marewski, J. N., & Schooler, L. J. (2011). Cognitive niches: An ecological model of strategy selection. *Psychological Review, 118*(3), 393–437. https://doi.org/10.1037/a0024143

Muldoon, K., Simms, V., Towse, J., Menzies, V., & Yue, G. (2011). Cross-cultural comparisons of 5-year-olds' estimating and mathematical ability. *Journal of Cross-Cultural Psychology, 42*(4), 669–681. https://doi.org/10.1177/0022022111406035

National Mathematics Advisory Panel. (2008). *Foundations for success: The final report of the National Mathematics Advisory Panel*. US Department of Education.

Newton, K. J., Lange, K., & Booth, J. L. (2019). Mathematical flexibility: Aspects of a continuum and the role of prior knowledge. *The Journal of Experimental Education, 88*(4), 503–515. https://doi.org/10.1080/00220973.2019.1586629

Nistal, A. A., Van Dooren, W., & Verschaffel, L. (2014). Improving students' representational flexibility in linear-function problems: An intervention. *Educational Psychology, 34*(6), 763–786. https://doi.org/10.1080/01443410.2013.785064

O'Sullivan, J., & Pressley, M. (1984). Completeness of instruction and strategy transfer. *Journal of Experimental Child Psychology, 38*(2), 275–288. https://doi.org/10.1016/0022-0965(84)90126-7
Oikarainen, V. (2013). *A model to analyse algebraic tasks solved by students: A comparative study from Finland and Norway* [Master's thesis]. University of Agder.
Pehkonen, E. (2008). *Problem solving in mathematics education in Finland*. Paper presented at the Symposium on the Occasion of the 100th Anniversary of ICMI (International Commission on Mathematical Instruction), Rome.
Ramirez, G., Chang, H., Maloney, E. A., Levine, S. C., & Beilock, S. L. (2016). On the relationship between math anxiety and math achievement in early elementary school: The role of problem solving strategies. *Journal of Experimental Child Psychology, 141*, 83–100. https://doi.org/10.1016/j.jecp.2015.07.014
Reinhardtsen, J., & Givvin, K. B. (2019). The fifth lesson: Students' responses to a patterning task across the four countries. In C. Kilhamn & R. Säljö (Eds.), *Encountering algebra* (pp. 165–234). Springer. https://doi.org/10.1007/978-3-030-17577-1
Rico, L. (1995). Errors and difficulties in mathematics learning [Errores y dificultades en el aprendizaje de las matemáticas]. In J. Kilpatrick, P. Gómez, & L. Rico (Eds.), *Educación matemática. Errores y dificultades de los estudiantes. Resolución de problemas. Evaluación. Historia* (pp. 69–108). Una Empresa Docente.
Rittle-Johnson, B., & Star, J. R. (2007). Does comparing solution methods facilitate conceptual and procedural knowledge?: An experimental study on learning to solve equations. *Journal of Educational Psychology, 99*(3), 561–574. https://doi.org/10.1037/0022-0663.99.3.561
Robitaille, D. F., & Travers, K. J. (1992). International studies of achievement in mathematics. In D. A. Grouws (Ed.), *Handbook of research on mathematics teaching and learning* (pp. 687–709). National Council of Teachers of Mathematics.
Röj-Lindberg, A.-S., & Partanen, A.-M. (2019). Learning to solve equations in three Swedish-speaking classrooms in Finland. In C. Kilhamn & R. Säljö (Eds.), *Encountering Algebra* (pp. 111–138). Springer. https://doi.org/10.1007/978-3-030-17577-1
Sawyer, R. J., Graham, S., & Harris, K. R. (1992). Direct teaching, strategy instruction, and strategy instruction with explicit self-regulation: Effects on the composition skills and self-efficacy of students with learning disabilities. *Journal of Educational Psychology, 84*(3), 340–352. https://doi.org/10.1037/0022-0663.84.3.340
Schneider, M., Rittle-Johnson, B., & Star, J. R. (2011). Relations among conceptual knowledge, procedural knowledge, and procedural flexibility in two samples differing in prior knowledge. *Developmental Psychology, 47*(6), 1525–1538. https://doi.org/10.1037/a0024997
Shaw, S. T., Pogossian, A. A., & Ramirez, G. (2020). The mathematical flexibility of college students: The role of cognitive and affective factors. *British Journal of Educational Psychology, 90*(4), 981–996. https://doi.org/10.1111/bjep.12340
Shrager, J., & Siegler, R. S. (1998). SCADS: A model of children's strategy choices and strategy discoveries. *Psychological Science, 9*(5), 405–410. https://doi.org/10.1111/1467-9280.00076
Siegler, R., & Araya, R. (2005). A computational model of conscious and unconscious strategy discovery. In R. V. Kail (Ed.), *Advances in child development and behavior* (Vol. 33, pp. 1–42). JAI. https://doi.org/10.1016/S0065-2407(05)80003-5
Siegler, R. S., & Lemaire, P. (1997). Older and younger adults' strategy choices in multiplication: Testing predictions of ASCM using the choice/no-choice method. *Journal of Experimental Psychology: General, 126*(1), 71–92. https://doi.org/10.1037/0096-3445.126.1.71
Siegler, R. S., & Shipley, C. (1995). Variation, selection, and cognitive change. In T. Simon & G. Halford (Eds.), *Developing cognitive competence: New approaches to process modeling* (pp. 31–76). Lawrence Erlbaum Associates.
Singapore Ministry of Education. (2006). *Secondary mathematics syllabuses*. Retrieved 2006 from https://www.moe.gov.sg/education/syllabuses/sciences/files/maths-secondary.pdf
Socas, M. M., Ruano, R. M., & Domínguez, J. H. (2016). Análisis Didáctico del proceso matemático de Modelización en alumnos de Secundaria. *Avances de Investigación en Educación Matemática, 9*, 21–41. https://doi.org/10.35763/aiem.v0i9.146
Sood, S., & Forehand, M. (2005). On self-referencing differences in judgment and choice. *Organizational Behavior and Human Decision Processes, 98*(2), 144–154. https://doi.org/10.1016/j.obhdp.2005.05.005
Star, J. R., & Madnani, J. (2004). Which way is "best"? Students' conceptions of optimal strategies for solving equations. *Proceedings of the twenty-sixth annual meeting of the North American*

Chapter of the International Group for the Psychology of Mathematics Education (Vol. 2, pp. 483–490). OISE/UT.

Star, J. R., Newton, K., Pollack, C., Kokka, K., Rittle-Johnson, B., & Durkin, K. (2015a). Student, teacher, and instructional characteristics related to students' gains in flexibility. *Contemporary Educational Psychology, 41*, 198–208. https://doi.org/10.1016/j.cedpsych.2015.03.001

Star, J. R., & Newton, K. J. (2009). The nature and development of experts' strategy flexibility for solving equations. *ZDM Mathematics Education, 41*(5), 557–567. https://doi.org/10.1007/s11858-009-0185-5

Star, J. R., Pollack, C., Durkin, K., Rittle-Johnson, B., Lynch, K., Newton, K., & Gogolen, C. (2015b). Learning from comparison in algebra. *Contemporary Educational Psychology, 40*, 41–54. https://doi.org/10.1016/j.cedpsych.2014.05.005

Star, J. R., & Rittle-Johnson, B. (2008). Flexibility in problem solving: The case of equation solving. *Learning and Instruction, 18*(6), 565–579. https://doi.org/10.1016/j.learninstruc.2007.09.018

Star, J. R., & Seifert, C. (2006). The development of flexibility in equation solving. *Contemporary Educational Psychology, 31*(3), 280–300. https://doi.org/10.1016/j.cedpsych.2005.08.001

Star, J. R., Tuomela, D., Prieto, N. J., Hästö, P., Palkki, R., Abánades, M. Á., et al. (2022). Exploring students' procedural flexibility in three countries. *International Journal of STEM Education, 9*(1), 1–18. https://doi.org/10.1186/s40594-021-00322-y

Threlfall, J. (2009). Strategies and flexibility in mental calculation. *ZDM Mathematics Education, 41*(5), 541–555. https://doi.org/10.1007/s11858-009-0195-3

Tian, J., Braithwaite, D. W., & Siegler, R. S. (2021). Distributions of textbook problems predict student learning: Data from decimal arithmetic. *Journal of Educational Psychology, 113*(3), 516–529. https://doi.org/10.1037/edu0000618

Torbeyns, J., & Verschaffel, L. (2013). Efficient and flexible strategy use on multi-digit sums: A choice/no-choice study. *Research in Mathematics Education, 15*(2), 129–140. https://doi.org/10.1080/14794802.2013.797745

Torbeyns, J., Verschaffel, L., & Ghesquiere, P. (2004). Strategy development in children with mathematical disabilities: Insights from the choice/no-choice method and the chronological-age/ability-level—match design. *Journal of Learning Disabilities, 37*(2), 119–131. https://doi.org/10.1177/00222194040370020301

Vega-Castro, D., Molina, M., & Castro, E. (2012). Sentido estructural de estudiantes de bachillerato en tareas de simplificación de fracciones algebraicas que involucran igualdades notables. *Revista Latinoamericana de Investigación en Matemática Educativa, 15*(2), 233–258.

Verschaffel, L., Luwel, K., Torbeyns, J., & Van Dooren, W. (2009). Conceptualizing, investigating, and enhancing adaptive expertise in elementary mathematics education. *European Journal of Psychology of Education, 24*(3), 335–359. https://doi.org/10.1007/Bf03174765

Verschaffel, L., Luwel, K., Torbeyns, J., & Van Dooren, W. (2011). Analyzing and developing strategy flexibility in mathematics education. In J. Elen, E. Stahl, R. Bromme, & G. Clarebout (Eds.), *Links between beliefs and cognitive flexibility* (pp. 175–197). Springer. https://doi.org/10.1007/978-94-007-1793-0_10

Vicente, S., Sánchez, R., & Verschaffel, L. (2020). Word problem solving approaches in mathematics textbooks: A comparison between Singapore and Spain. *European Journal of Psychology of Education, 35*(3), 567–587. https://doi.org/10.1007/s10212-019-00447-3

Wang, J., Liu, R.-D., Star, J. R., Liu, Y., & Zhen, R. (2019). The moderating effect of regulatory focus in the relationship between potential flexibility and practical flexibility. *Contemporary Educational Psychology, 56*, 218–227. https://doi.org/10.1016/j.cedpsych.2019.01.013

West, P. M., Huber, J., & Min, K. S. (2004). Altering experienced utility: The impact of story writing and self-referencing on preferences. *Journal of Consumer Research, 31*(3), 623–630. https://doi.org/10.1086/425097

Woodward, J., Beckmann, S., Driscoll, M., Franke, M., Herzig, P., Jitendra, A., Koedinger, K. R., & Ogbuehi, P. (2012). *Improving mathematical problem solving in grades 4 to 8: A practice guide (NCEE 2012-4055)*. National Center for Education Evaluation and Regional Assistance.

Xu, C., Wells, E., LeFevre, J.-A., & Imbo, I. (2014). Strategic flexibility in computational estimation for Chinese- and Canadian-educated adults. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 40*(5), 1481–1497. https://doi.org/10.1037/a0037346

Xu, L., Liu, R.-D., Star, J. R., Wang, J., Liu, Y., & Zhen, R. (2017). Measures of potential flexibility and practical flexibility in equation solving. *Frontiers in Psychology, 8*, 1368. https://doi.org/10.3389/fpsyg.2017.01368

Yakes, C., & Star, J. R. (2011). Using comparison to develop flexibility for teaching algebra. *Journal of Mathematics Teacher Education, 14*(3), 175–191. https://doi.org/10.1007/s10857-009-9131-2

Yang, D.-C., & Sianturi, I. A. J. (2020). Analysis of algebraic problems intended for elementary graders in Finland, Indonesia, Malaysia, Singapore, and Taiwan. *Educational Studies*, *48*(1), 75–97. https://doi.org/10.1080/03055698.2020.1740977

Yeap, B.-H., Ferrucci, B. J., & Carter, J. A. (2006). Comparative study of arithmetic problems in Singaporean and American mathematics textbooks. In F. K. S. Leung, K. D. Graf, & F. J. Lopez-Real (Eds.), *Mathematics education in different cultural traditions—A comparative study of East Asia and the West* (pp. 213–225). Springer. https://doi.org/10.1007/0-387-29723-5_13

Zohar, A. (2012). Explicit teaching of metastrategic knowledge: Definitions, students’ learning, and teachers’ professional development. In A. Zohar & Y. J. Dori (Eds.), *Metacognition in science education: Trends in current research* (pp. 197–223). Springer Netherlands.https://doi.org/10.1007/978-94-007-2132-6_9

Zohar, A., & David, A. B. (2008). Explicit teaching of meta-strategic knowledge in authentic classroom situations. *Metacognition and Learning, 3*(1), 59–82. https://doi.org/10.1007/s11409-007-9019-4

Zohar, A., & Peled, B. (2008). The effects of explicit teaching of metastrategic knowledge on low-and high-achieving students. *Learning and Instruction, 18*(4), 337–353. https://doi.org/10.1016/j.learninstruc.2007.07.001

Authors and Affiliations

Ronghuan Jiang[1] · Jon R. Star[2] · Peter Hästö[3] · Lijia Li[2] · Ru-de Liu[4] · Dimitri Tuomela[3] · Nuria Joglar Prieto[5] · Riikka Palkki[3] · Miguel Á. Abánades[6] · Johanna Pejlare[7]

1 School of Psychology, Shenzhen University, Shenzhen, China

2 Harvard Graduate School of Education, Cambridge, MA, USA

3 University of Oulu, Oulu, Finland

4 Faculty of Psychology, Beijing Normal University, Room 1412, Main building, Beijing 100875, China

5 Universidad Complutense de Madrid, Madrid, Spain

6 Universidad Rey Juan Carlos, Madrid, Spain

7 Chalmers University, Gothenburg, Sweden