

RESEARCH

Open Access



Exploring students' procedural flexibility in three countries

Jon R. Star^{1*} , Dimitri Tuomela², Nuria Joglar-Prieto³, Peter Hästö², Riikka Palkki², Miguel Á. Abánades⁴, Johanna Pejlare⁵, R. H. Jiang⁶, Lijia Li¹ and Ru-De Liu⁶

Abstract

Background: In this cross-national study, Spanish, Finnish, and Swedish middle and high school students' procedural flexibility was examined, with the specific intent of determining whether and how students' equation-solving accuracy and flexibility varied by country, age, and/or academic track. The 791 student participants were asked to solve twelve linear equations, provide multiple strategies for each equation, and select the best strategy from among their own strategies.

Results: Our results indicate that knowledge and use of the standard algorithm for solving linear equations is quite widespread across students in all three countries, but that there exists substantial within-country variation as well as between-country variation in students' reliance on standard vs. situationally appropriate strategies. In addition, we found correlations between equation-solving accuracy and students' flexibility in all three countries but to different degrees.

Conclusions: Although it is increasingly recognized as an important construct of interest, there are many aspects of mathematical flexibility that are not well-understood. Particularly lacking in the literature on flexibility are studies that explore similarities and differences in students' repertoire of strategies for solving algebra problems across countries with different educational systems and curricula. This study yielded important insights about flexibility and can push the field to explore the extent that within- and between-country differences in flexibility can be linked to differences in countries' educational systems, teaching practices, and/or cultural norms around mathematics teaching and learning.

Keywords: Procedural flexibility, Algebra, Linear equation solving, Cross-national

Introduction

Flexibility in the use of mathematical procedures—or procedural flexibility—has emerged as an important outcome in educational policy and practice (National Council of Teachers of Mathematics [NCTM], 2014; National Research Council, 2001; Star, 2005). As noted in a position paper on procedural fluency from the National Council of Teachers of Mathematics (NCTM, 2014): “All students need to have a deep and flexible

knowledge of a variety of procedures, along with an ability to make critical judgments about which procedures or strategies are appropriate for use in particular situations.” Researchers have begun to investigate procedural flexibility, in mathematical domains including arithmetic (Blöte et al., 2001; Shaw et al., 2020; Torbeyns et al., 2009), computational estimation (Star & Rittle-Johnson, 2009), algebra (e.g., Rittle-Johnson & Star, 2007), linear algebra (Maciejewski & Star, 2019) and calculus (Maciejewski & Star, 2016); and among

*Correspondence: jon_star@harvard.edu

¹ Harvard Graduate School of Education, 442 Gutman Library, Cambridge, MA 02138, USA

Full list of author information is available at the end of the article

American (Rittle-Johnson & Star, 2009) and international (Hästö et al., 2019; Joglar et al., 2018; Xu et al., 2017) students.¹ Within this literature, procedural flexibility is defined as knowledge of multiple strategies and the ability to select the most appropriate strategy for a given problem and problem-solving circumstances (e.g., Star, 2005).

Algebra, particularly linear equation solving, has proved to be a particularly productive content area for research on mathematical flexibility (e.g., Huntley et al., 2007; Star et al., 2015). In many countries around the world, students are introduced to linear equation solving at roughly the same age (12 or 13 years). For linear equations, such as $3(x + 1) = 15$, a standard algorithm exists (the first step of which is to distribute the 3; Buchbinder et al., 2015; Star & Seifert, 2006), yet there are other alternative strategies that are arguably more optimal for this particular equation, such as dividing both sides of the equation by 3 as a first step. We refer to strategies that are arguably better than a standard algorithm as “situationally appropriate.” With this label, we acknowledge that it is rarely the case that one strategy is always the best but rather that what it means for a strategy to be better than another is often dependent on the solver’s beliefs and goals as well as the particular tasks to be solved (Verschaffel et al., 2009). Findings from the literature suggest that students do not always select a situationally appropriate strategy for a given equation, even though they have knowledge of multiple strategies (Newton et al., 2010). Furthermore, even gifted students or mathematics content experts who have knowledge of multiple strategies do not always choose a situationally appropriate strategy (Dover & Shore, 1991; Star & Newton, 2009). There appears to be a difference between what one knows and what one decides to use spontaneously during problem solving (Dover & Shore, 1991; Flavell & Wohlwill, 1969).

As mentioned above, core to procedural flexibility is the ability to select a situationally appropriate strategy for a given problem, from among a solver’s knowledge of multiple problem-solving strategies. Yet it is also the case that identifying such a strategy that is most appropriate can be quite subtle and nuanced (Hatano & Inagaki, 1984; Verschaffel et al., 2009). In some cases, the most efficient strategy—i.e., the strategy with the fewest steps, or the strategy that can be executed the

quickest—may be considered situationally appropriate. However, it might also be the case that the strategy that can be executed most reliably and without error could be viewed as situationally appropriate. More generally, within the discipline of mathematics, mathematicians tend to believe that a situationally appropriate strategy is the one that is most elegant, despite the fact that it is often difficult to objectively define elegance (Hardy, 1940).

The relationship between flexibility and solving accuracy

In addition to the identification of procedural flexibility as a learning outcome (NCTM, 2014; Star, 2005), there is also an expectation in the literature that flexibility is related to accuracy (e.g., Rittle-Johnson & Star, 2007). In other words, increased procedural flexibility is believed to support increased problem-solving accuracy. There is some indirect evidence in support of this assumption. For example, Rittle-Johnson and colleagues (Rittle-Johnson & Star, 2007, 2009; Rittle-Johnson et al., 2009, 2012; Star & Rittle-Johnson, 2008) found that a contrasting cases intervention in the domain of linear equation solving generally led to gains in both procedural flexibility and accuracy. Yet at the same time, Xu and colleagues (Xu et al., 2017) found that procedural flexibility was not strongly correlated with accuracy scores, in that students who were very accurate equation solvers did not consistently demonstrate high flexibility.

There are other recent studies that have explored possible relationships between strategy flexibility and problem-solving accuracy in different mathematical domains and among students from a variety of age levels. For example, Carr and Taasoobshirazi (2017) found that early variability in primary school mathematics strategy use was linked to positive learning outcomes in later years. McMullen and colleagues (McMullen et al., 2017) found that students’ strategies for working with rational numbers predicted later pre-algebra skills. Similarly, Levav-Waynberg and Leikin (2012) found a relationship between geometrical knowledge and the strategies used by students who engaged with tasks for which there were multiple possible solution methods. In addition, Lemaire and Siegler (1995) found that improved adaptivity in strategy use was one explanation for increased speed and accuracy in multiplication tasks for French 2nd graders. Taken together, these studies suggest that there is a relationship between the strategies that students use and their accuracy in problem solving, although (as noted above) this evidence is not conclusive with respect to procedural flexibility.

¹ Although less widely studied, flexibility also plays a key role in other STEM disciplines, including spatial thinking (Uttal, Miller, & Newcome, 2013), electronics (Pirttimaa, Husu, & Metsärinne, 2017), fingerprint analyses (Mustonen & Hakkarainen, 2015), and biomedical engineering ethics (Martin, Rayne, Kemp, Hart, & Diller, 2005).

The relationship between flexibility and mathematical expertise

Perhaps a corollary to the presumed relationship between procedural flexibility and accuracy is the expectation that older students (or those with greater mathematical expertise) will also be higher in flexibility. The correlation between flexibility and age and/or mathematical expertise has been suggested in other mathematical domains, such as computational estimation (Dowker, 1992) or in qualitative studies of struggling students (Lynch & Star, 2014). In addition, Chinese 8th grade students who showed high accuracy in equation solving—successfully solving an average of 10.84 out of 12 linear equations, or 90.3%—also exhibited relatively high levels of procedural flexibility (Liu et al., 2018). However, at the same time, other studies have found that both younger and older students possess relatively low levels of flexibility, with little or no growth as student's age and presumably become more mathematically knowledgeable. For example, Star and Seifert (2006) found that US 6th graders exhibited flexibility in only about 9% of post-test problems, despite arriving at correct answers on 77% of problems. Similarly, Lewis (1981) found that US undergraduates exhibited flexibility at a comparable rate (9–12% of problems), with mathematicians only somewhat better (about 20% of problems). (Recent work suggests a possible link between flexibility and confidence in mathematics, which could begin to help explain the differences in flexibility found in past studies with US students and Chinese students (Maciejewski, 2020).) However, note that studies specifically designed to examine the relationship between procedural flexibility and age—i.e., involving students at multiple grade levels, either via a cross-sectional or longitudinal design—are quite rare in the literature on flexibility. As a result, the extent of the relationship between procedural flexibility, age, and mathematical expertise is still largely unexplored.

Cross-cultural differences in flexibility

Furthermore, the current literature on flexibility also suggests that there may be substantial cross-cultural differences in both problem-solving accuracy as well as procedural flexibility. As noted above, both Liu and colleagues (Liu et al., 2018) and Xu and colleagues (Xu et al., 2017) found that Chinese middle school students exhibited very high accuracy and moderate levels of procedural flexibility when solving linear equations. Yet studies with US middle school students (e.g., Rittle-Johnson & Star, 2007; Star & Seifert, 2006) found both accuracy and procedural flexibility rates to be substantially lower than what was found among the Chinese students. Although cross-cultural studies of mathematical problem solving exist in the literature (e.g., Cai, 2000; Cai & Hwang, 2002;

Chen & Stevenson, 1995), no prior cross-cultural studies have examined students' procedural flexibility.

There are several reasons to believe that procedural flexibility may indeed differ cross-culturally. In particular, prior studies have consistently found that students from different countries demonstrated different types of mathematical thinking when solving problems (Cai, 2004; Jiang et al., 2014, 2017). For example, Cai (2000, 2004) found that American students exhibited greater diversity of thinking and also used more uncommon strategies when solving mathematics problems as compared to their Chinese peers, although Chinese students' accuracy was higher. Similarly, a recent study comparing Spanish and Chinese 4th to 8th grade students' performance on addition problems and proportion problems (Jiang et al., 2017) found differences between the reasoning used by Chinese and Spanish students. Relatedly, Gorgorió et al. (2018) found differences between Spanish and Finnish university students' mathematical knowledge and strategies upon entry to a primary teacher education program. More generally, curriculum has been found to influence mathematics learning in international comparative studies of mathematics achievement (Martin et al., 2008; Mullis et al., 2012) suggesting that curricular differences between countries would also influence students' procedural flexibility. Yet at the same time, other studies have found a striking degree of uniformity amongst students' mathematical understandings (or lack thereof) and strategies in many countries. For example, previous studies found that students from several different countries showed a strong tendency to overuse a certain type of proportional reasoning strategy to solve a set of problems and were not capable of switching amongst different strategies based on the types of problem (De Bock et al., 2002, 2007; Fernández et al., 2012; Li et al., 2014). Similarly, various misconceptions related to the use of negative signs, equals signs, variables, and fractions appear to be prevalent in a variety of countries (Booth et al., 2014; Bush & Karp, 2013). As a result, whether (and the degree that) there are cross-cultural differences in students' procedural flexibility is an open question, the answer to which will be informative to efforts in many countries to promote the development of procedural flexibility as one instructional outcome.

The present study

We study the development of procedural flexibility (following Star's definition; Star, 2005) in the context of solving linear equations in secondary mathematics education. In particular, the present study takes advantage of the recently developed validated assessment of procedural flexibility in equation solving, the Tri-Phase Flexibility Assessment (Xu et al., 2017) (described in more

depth below). Using convenience samples of students from three countries (Spain, Finland, and Sweden), a cross sectional design (assessing both middle school and high school students), and working with students in both basic and advanced tracks in mathematics, we explore the following research questions.

- RQ1 *How does accuracy in linear equation solving vary by age (middle school vs. high school), by academic track (basic vs. advanced track), and by country (Spain, Finland, and Sweden)?* As noted above, there is mixed evidence in the prior literature as to the presence of cross-cultural differences in mathematical proficiency. We seek to add to the existing cross-national results found in studies such as PISA and TIMSS to more specifically document cross-cultural differences in mathematical problem-solving within the particular domain of linear equation solving.
- RQ2 *How does procedural flexibility in linear equation solving vary by age, by academic track, and by country?* As noted above, previous research has not addressed how country, academic track or age are related to students' procedural flexibility. By investigating students' procedural flexibility in various contexts, we aim to identify factors that inhibit or enhance flexibility, which in turn may (in future research) shed light on how procedural flexibility might develop over time.
- RQ3 *To what extent are procedural flexibility and accuracy related? Does this relationship vary by age, academic track, and/or country?* Although the prior literature on procedural flexibility implicitly assumes that procedural flexibility is related to accurately solving mathematical tasks, there is little direct evidence in support of this relationship—and some of this evidence is mixed. While the development of procedural flexibility is itself a reasonable instructional goal, procedural flexibility without accuracy is certainly not optimal and could be counterproductive. In particular, when, where, and under which circumstances do both procedural flexibility and procedural accuracy develop?

Method

Participants

A total of 791 middle school and high school students from Finland, Sweden, and Spain participated in the study. The convenience samples of schools and students

were recruited by members of the research team who were local to each country. The schools were diverse in terms of geographic location and size. It is also worth noting that the public educational systems in Finland, Sweden and Spain are mostly composed of schools whose students are relatively homogenous in terms of ethnicity and socio-economic status, at least as compared to the United States (e.g., Boli, 2014; Sahlberg, 2014).

These three countries were selected both for convenience but also because they have (among western European countries) an interesting mix of both similarity and variation among various educational and mathematics educational indicators, suggesting that cross-national comparisons could be productive. For example, all three countries differ to some degree on indicators, such as 2015 PISA scores (511, 494 and 486, for Finland, Sweden, and Spain, respectively), 2017 public expenditure in education (5.7% of GDP, 6.8% of GDP, and 4.0% of GDP, respectively), and in the proportion of 15 years who are underachieving in mathematics (13.6%, 20.8%, and 22.2%, respectively).² Furthermore, while students from these three countries can be considered to be educated in culturally similar western European contexts, the literature suggests that there may be interesting differences in their typical mathematics teaching and learning environments. In particular, there is some evidence that the secondary mathematics teaching and learning environment in Spain may be quite different than in Finland and Sweden, two countries that may have many similarities in views of how math is taught. For example, researchers have found the Spanish mathematics curriculum (especially in secondary school) to be very traditionally teacher centered, with emphasis on the routine application of rules with a focus on accuracy and speed, and with classwork, assessments, and homework involving a great deal of timed practice (e.g., González-Astudillo & Sierra-Vázquez, 2004). In contrast, in Finland and Sweden, the literature suggests that one more typically finds a de-emphasis on timed drills (Hemmi & Krzywacki, 2014; Pehkonen et al., 2018). As a result, while primarily driven by convenience, our investigation of students in Finland, Sweden, and Spain provides research contexts that have interesting similarities (between Finland and Sweden) as well as differences (in Spain) to allow for a fruitful exploration of how procedural flexibility might vary across countries.

In Finland, 257 students from 8 schools participated in the study, 93 of whom were in 8th grade (middle school) and the remaining 164 in 11th grade (high school). The eight schools were selected to represent typical schools in Finland; note that demographic information is not

² <https://ec.europa.eu/eurostat/web/education-and-training/eu-benchmarks>.

generally collected in Finland, since schools are very homogeneous in terms of student characteristics. Of the high school students, 103 were in the advanced mathematics track, while 61 were in the basic track. Finnish children start mandatory schooling at age 7 and attend compulsory school until the end of 9th grade, using a standard mathematics curriculum. In high school students choose between two tracks in mathematics: advanced and basic. Attendance in grades 10–12 is not compulsory, and approximately 50% of students choose to attend.

In Sweden, 288 students from 6 schools participated in the study, 87 of whom were in the 9th grade (middle school) and the remaining 201 in the 10th grade (high school). The six schools were selected to represent typical public schools in Sweden. Two schools were located in a larger city and the remaining four schools were in four different smaller cities. Based on statistics provided by the Swedish National Agency for Education,³ it is estimated that around 25% of the participating students came from a foreign background (i.e., were either born outside of Sweden or both parents were born outside of Sweden), which is similar to the national average. Of the high school students, 181 were in the advanced track, while 20 were in the basic track. Swedish children start compulsory schooling at age 7 and attend compulsory school until the end of 9th grade, using a standard mathematics curriculum. Attendance in grades 10–12 is not compulsory, and over 95% of students choose to attend.

In Spain, 246 students from 5 schools participated in the study, 164 of whom were in 8th grade (middle school) and the remaining 82 were in 11th grade (high school). Participating schools were similar to typical urban public schools in Spain, in that students were predominantly middle class and of Spanish background with gender parity. Regarding social background, the five participating schools were from two different geographic regions. Four schools were located in a region with an average immigration ratio of 9% (nationally, approximately 8.7% of students in Spain are foreign immigrants) and one school in a region with 4% of immigrant students. Of the high school students, 47 were in the advanced track in mathematics, while 35 were in the basic track. Spanish children start mandatory school at age 6 and attend compulsory school until the end of 10th grade, using a standard mathematics curriculum. Attendance in grades 11 and 12 is not compulsory; approximately 70% of Spanish students choose to attend.

Note that in each of the three countries, multi-step linear equation solving is first introduced in 7th grade.

In addition, although we used a convenience sample of schools that was not randomly or representatively sampled, research team members local to each country reported that the schools selected for participation in this study were considered to be very typical and not exceptional along any dimension, including number of students, proportion of males and females, proportion of students with a foreign background, proportion of students whose parents have college degrees, etc. Finally, despite the different structures of the education systems, we estimate that in each country approximately 20% of the age cohort attend what we call advanced mathematics in high school.

For each country, the middle school sample included students who would ultimately go on to high school in either the advanced mathematics track or the basic mathematics track, as well as those who would not continue their schooling beyond what was mandatory in each country. In the interest of facilitating comparison of results between middle school students and high school students who were in advanced mathematics, we identified the top 20% of middle school students in each country's sample using overall accuracy scores. Note that this "advanced middle school" category is merely a subset of the sample of students from middle school, given that (unlike in the high schools) there is no tracking of students in middle school in any of the participating countries. The advanced middle school cohort is a reasonable group to compare with the advanced high school students within and between countries, since the latter group consists, roughly speaking, of the top 20% of students in the age cohort.⁴

Because of the different educational systems being examined, it was not always possible to compare identical samples; for example, the middle school students in the present study were in 8th grade in Spain and Finland but in 9th grade in Sweden. The distribution of students in each country at each level varied considerably. For example, in Spain only 47 of the 246 students (19%) were in the advanced mathematics track in high school, while in Finland and Sweden, these proportions were 40% (103 of 258) and 63% (181 of 288), respectively. We have chosen measures and made comparisons in such a way as to minimize the effect of these differences; specifically, averages over countries are calculated by assigning equal weight to each country, not each student.

³ <https://www.skolverket.se/skolutveckling/statistik>.

⁴ Our data do not allow us to precisely determine the extent that this top 20% of middle school students in each country's sample did indeed go on to take more advanced mathematics classes in subsequent years. However, anecdotal reports in each country suggests that it is extremely likely that this would have been the case.

Table 1 Assessment problems, with standard algorithms and situationally appropriate methods

#	Problem	Standard algorithm	Innovative strategy
1	$4(x - 2) = 24$	Begin by distributing the parentheses	Divide a constant to both sides before distributing
2	$3(x + 0.69) = 15$		
3	$4\left(x + \frac{3}{5}\right) = 12$		
4	$4(x + 6) + 3(x + 6) = 21$	Begin by distributing the parentheses	Change in variable – combine
5	$5\left(x + \frac{3}{7}\right) + 3\left(x + \frac{3}{7}\right) = 16$		
6	$2(x - 0.31) + 3(x - 0.31) = 15$		
7	$8(x - 5) = 3(x - 5) + 20$	Begin by distributing the parentheses	Change in variable – subtract from both
8	$8\left(x - \frac{2}{5}\right) - 11 = 6\left(x - \frac{2}{5}\right)$		
9	$5(x + 0.6) + 3x = 5(x + 0.6) + 7$		
10	$\frac{2x-6}{2} + \frac{6x-18}{3} = 5$	Begin by obtaining a common denominator for the two expressions	Reducing each fraction before combining
11	$\frac{x+3}{3} + \frac{3x+9}{9} = 1$		
12	$\frac{5x+5}{5} + \frac{6x+6}{6} = 6$		

Information about students’ gender, race, and socio-economic status was not collected. We discuss this issue below as an important limitation of this study.

Measures and procedures

All students completed a translation⁵ of the Tri-Phase Flexibility Assessment (Xu et al., 2017), which was designed to assess students’ procedural flexibility, potential flexibility, and spontaneous flexibility in linear equation solving (we define these constructs below). The assessment consists of a test, instructions on how it is to be administered, and specifications on how to code the data. Note that Xu and colleagues (Xu et al., 2017) explored the psychometric properties of the Tri-Phase Flexibility Assessment and found it to be valid and reliable; in particular, Xu and colleagues found this instrument to have high internal consistency (Cronbach’s alpha) for measuring flexibility and spontaneous flexibility, as well as good composite reliability and convergent validity for these constructs. As a result, we have followed their procedure as closely as possible in the use of and administration of the assessment.

The test consisted of 12 linear equations to be solved in three phases, which were implemented in the present study identically to the procedures used by Xu et al. (2017). In Phase 1, students were asked to solve each of the 12 problems as quickly and as accurately as possible.

Students were instructed to write their solutions in a box that was clearly marked. Students were provided 15 min for completing this phase. In Phase 2, students returned to the same 12 problems and were asked to generate as many different solutions for each problem as possible (e.g., in Finnish: *kirjoittaa jokaiseen tehtävään niin monta eri ratkaisua kuin mahdollista*; in Spanish: *resolver la ecuación de todas las formas diferentes posibles que podáis*; in Swedish: *lös varje uppgift på så många olika sätt som möjligt*); the test provided space for up to 5 different strategies. Note that we did not provide any further explanation or examples about what constituted a “different” strategy, as we were interested to see how students themselves interpreted this request. Students were provided 20 min for this phase.

Finally, in Phase 3, students were asked to return to the 12 problems and to circle the strategy for each problem (from among the multiple strategies that they produced) that they considered to be the best (e.g., in Finnish: *paras*; in Spanish: *la major*; in Swedish: *bäst*). As with the word “different,” we did not provide explanation to guide students in their selection of the best strategy, as we wanted to see how students themselves interpreted this request. Students were given 5 min for this phase.

The assessment included four types of linear equations, with three of each type (see Table 1). Note that, for the first three problem types, the three equations for each type had the same structure but differed in terms of the coefficients and constants in each problem: The first problem of each type (Problems 1, 4, and 7) had integer coefficients and constants. Of the other two problems, one used fractions (Problems 3, 5, and 8) and one used decimals (Problems 2, 6, and 9).

⁵ To ensure that the translations of this instrument were accurate, we began by working with the native English-speaking and native Chinese-speaking authors of the original Xu et al. assessment to create an English version of the instrument. This version was then translated into Finnish, Spanish, and Swedish by native-speaking collaborators in each country and then back-translated and checked for accuracy by the first author, who is a native English speaker.

All equations could be solved with the following standard algorithm that is taught as an explicit component of the mathematics curriculum in Finland, Sweden, and Spain and that has been used in prior research on linear equation solving (e.g., Buchbinder et al., 2015; Rittle-Johnson & Star, 2007; Star & Seifert, 2006): First, use the distributive property to ‘clear’ parentheses. Second, combine like variable and constant terms on the left and right sides of the equation. Third, add or subtract to both sides, putting variable terms on the left side and constant terms on the right side of the equation, resulting in an equation of the form $ax = b$. Finally, solve for the variable by dividing both sides by the coefficient of the variable term.

Each equation could also be solved by one or more situationally appropriate strategies, which was optimally matched to the structural features of the problem and resulted in fewer calculations to reach the solution (see Table 1). For the first three equation types, a situationally appropriate strategy involved treating the common parenthetical terms as a variable (i.e., a ‘change in variable’ strategy; Star & Seifert, 2006), either by dividing to both sides as a first step (Problems 1–3), combining like terms (Problems 4–6), or subtracting from both sides as a first step (Problems 7–9). For the fourth equation type, a situationally appropriate strategy involved simplifying each fractional term before combining (Problems 10–12).

Analysis

Each test was scored by one or two mathematics education professors or graduate students who were educated in the target country and also fluent in the language of instruction in that country. Scorers began by examining students’ solutions produced in Phase 1—i.e., those written in the space designated for each problem on the test to hold the solutions provided in Phase 1. Scorers then determined the *accuracy* of each attempt made during Phase 1, by evaluating whether a student arrived at the correct numerical answer by correct intermediate steps on the Phase 1 solution attempt for each problem. No partial credit was awarded, and thus the accuracy score for each student was an integer between 0 and 12. Scorers then examined all solution strategies produced during Phases 1 and 2 for each problem, coding each strategy as to whether it followed the standard algorithm described above, the situationally appropriate strategy illustrated in Table 1, or some other strategy. The assessment problems were designed so that the identification of a student’s strategy as situationally appropriate or standard could be usually determined by looking only at the initial steps of the student’s work. To this point, our coding was identical to that of Xu et al (2017).

Once coding of strategy type and accuracy were completed, students’ procedural flexibility, potential

flexibility, and spontaneous flexibility were calculated, using a procedure that was very similar to what has been used in prior studies with the Tri-Phase Flexibility Assessment (Xu et al., 2017).⁶ Each student was deemed *flexible* (*F*) on a given equation if they exhibited procedural flexibility on that problem, e.g., if they met the following three criteria:

- A. the student exhibited knowledge of the standard solution method for that equation,
- B. the student exhibited knowledge of a situationally appropriate solution method for that equation, and
- C. the student identified (in Phase Three) the situationally appropriate strategies as best for that equation.

The flexibility score for each student was an integer ranging from 0 to 12.

If a student provided the situationally appropriate strategy but failed to meet one of the other two criteria, they were deemed to show *potential flexibility* (*PF*) for that equation, indicating that their knowledge of the situationally appropriate strategy suggested that they may be on the verge of becoming procedurally flexible. The maximum potential flexibility score for each student was 12, although note that a student could either be flexible (*F*) or potentially flexible (*PF*) on a problem but not both.

Also note that there were two different types of potential flexibility, depending on which of the other flexibility criteria (A, B, or C; see above) was not satisfied for a given equation. *PF-AB* indicated that the student did not identify the situationally appropriate strategy as best among those that were generated, and *PF-BC* indicated that the student did not make use of the standard algorithm for that equation.⁷

Finally, for any student who demonstrated flexibility on a given equation, if a situationally appropriate strategy was used in the first solution attempt (i.e., in Phase 1 of

⁶ After engaging in multiple discussions with the authors of the Xu et al. (2017) paper, we made several changes to their scoring system and flexibility subconstruct nomenclature, which we (as well as the Xu et al. authors) believe to be an improvement in the assessment of flexibility and related subconstructs. First, what Xu et al. refer to as “practical flexibility,” we now refer to as “spontaneous flexibility” – which indicates whether a student used a situationally appropriate strategy in the first solution attempt. In contrast to Xu et al., we consider spontaneous flexibility to be a special and somewhat atypical outcome, rather than a desirable outcome for all students on all problems. Second, what Xu et al. refer to as “potential flexibility,” we now refer to as “flexibility”; our three criteria for flexibility are identical to Xu et al.’s three criteria for potential flexibility. We consider flexibility (as we define it here) to be the main outcome of interest, rather than an indicator that a student is approaching the main outcome of interest (as is the case in Xu et al.). And finally, we introduce a more nuanced notion of “potential flexibility” to capture indicators that flexibility is emerging for a student.

⁷ A third type of potential flexibility, *PF-AC*, while theoretically possible, cannot practically occur, as criterion C requires the presence of criterion B.

Table 2 Summary of key variables (mean and standard deviation) related to flexibility, for *all* students in Finland, Sweden and Spain

		<i>n</i>	<i>Acc</i>	<i>F</i>	<i>SF</i>	<i>PF</i>	<i>PF-AB</i>	<i>PF-BC</i>
Weighted averages over countries	Avg	791	5.4 (2.0)	1.8 (1.6)	0.5 (0.9)	0.6 (0.7)	0.6 (0.7)	0.0 (0.2)
	MS	344	3.1 (2.9)	0.5 (1.4)	0.1 (0.6)	0.3 (0.9)	0.3 (0.9)	0.0 (0.2)
	HS-B	116	5.1 (2.5)	1.4 (2.1)	0.2 (0.8)	0.6 (0.9)	0.6 (0.9)	0.0 (0.2)
	HS-A	331	8.0 (3.0)	3.6 (3.5)	1.3 (2.5)	0.9 (1.4)	0.8 (1.4)	0.1 (0.4)
Finland	Avg	257	4.5 (2.1)	1.7 (1.6)	0.6 (1.0)	0.5 (0.6)	0.4 (0.6)	0.0 (0.1)
	MS	93	1.6 (2.3)	0.5 (1.7)	0.2 (1.0)	0.1 (0.4)	0.1 (0.4)	0.0 (0.2)
	HS-B	61	3.9 (2.3)	0.4 (1.1)	0.0 (0.1)	0.4 (0.7)	0.3 (0.7)	0.0 (0.2)
	HS-A	103	7.9 (2.7)	4.3 (3.2)	1.5 (2.5)	0.9 (1.5)	0.8 (1.4)	0.1 (0.3)
Sweden	Avg	288	5.1 (1.8)	2.3 (1.7)	0.8 (1.1)	0.7 (0.6)	0.6 (0.5)	0.1 (0.2)
	MS	87	3.0 (2.5)	0.7 (1.6)	0.0 (0.3)	0.5 (1.0)	0.4 (0.9)	0.0 (0.3)
	HS-B	20	4.9 (1.5)	2.0 (1.7)	0.3 (0.6)	0.8 (0.9)	0.7 (0.9)	0.1 (0.2)
	HS-A	181	7.5 (3.4)	4.3 (3.6)	2.2 (2.9)	0.7 (1.0)	0.5 (0.9)	0.2 (0.5)
Spain	Avg	246	6.7 (1.8)	1.4 (1.5)	0.2 (0.5)	0.7 (0.8)	0.7 (0.8)	0.0 (0.0)
	MS	164	4.8 (2.8)	0.3 (0.9)	0.0 (0.2)	0.4 (1.2)	0.4 (1.2)	0.0 (0.1)
	HS-B	35	6.6 (2.7)	1.7 (2.8)	0.3 (1.1)	0.7 (1.1)	0.7 (1.1)	0.0 (0.0)
	HS-A	47	8.7 (2.6)	2.1 (3.2)	0.2 (1.2)	1.1 (1.7)	1.1 (1.8)	0.0 (0.0)

MS middle school; MS-A advanced middle school students; HS-A advanced mathematics track high school; HS-B basic track high school; Acc accuracy; F flexibility; SF spontaneous flexibility; PF potential flexibility

Table 3 Summary of key variables (mean and standard deviation) related to flexibility, for *high-achieving* students in Finland, Sweden and Spain

		<i>n</i>	<i>Acc</i>	<i>F</i>	<i>SF</i>	<i>PF</i>	<i>PF-AB</i>	<i>PF-BC</i>
Weighted averages over countries	MS-A	90	6.6 (2.4)	1.2 (2.3)	0.3 (1.2)	0.7 (1.5)	0.7 (1.5)	0.0 (0.2)
	HS-A	331	8.0 (3.0)	3.6 (3.5)	1.3 (2.5)	0.9 (1.4)	0.8 (1.4)	0.1 (0.4)
Finland	MS-A	24	4.8 (2.3)	1.5 (2.9)	0.6 (1.9)	0.3 (0.8)	0.2 (0.7)	0.1 (0.4)
	HS-A	103	7.9 (2.7)	4.3 (3.2)	1.5 (2.5)	0.9 (1.5)	0.8 (1.4)	0.1 (0.3)
Sweden	MS-A	22	6.5 (1.3)	1.3 (2.1)	0.2 (0.5)	1.0 (1.4)	1.0 (1.4)	0.0 (0.0)
	HS-A	181	7.5 (3.4)	4.3 (3.6)	2.2 (2.9)	0.7 (1.0)	0.5 (0.9)	0.2 (0.5)
Spain	MS-A	44	8.5 (1.6)	0.8 (1.5)	0.0 (0.3)	0.9 (1.9)	0.9 (1.9)	0.0 (0.0)
	HS-A	47	8.7 (2.6)	2.1 (3.2)	0.2 (1.2)	1.1 (1.7)	1.1 (1.8)	0.0 (0.0)

the assessment), this student was said to have *spontaneous flexibility (SF)* for that equation. The maximum possible spontaneous flexibility score for each student was again 12.

Results

We discuss results in terms of country-specific patterns of performance as well as several findings that compare performances across countries. Note that since our sample is not a representative one from each country, it is not particularly meaningful to draw conclusions from the absolute numbers in our results. Rather, we look for patterns, where multiple indicators are aligned—and for this reason, we mostly avoid statistical tests.

Tables 2, 3 shows students’ accuracy, flexibility, spontaneous flexibility and potential flexibility (including its two subtypes) for the Spanish, Swedish, and Finnish middle and high school students.

Procedural flexibility

Procedural flexibility was quite modest; across all groups of participants in the study, average demonstrated flexibility was 1.8 problems (out of 12, or 15% of equations; see Table 2). Swedish students had slightly higher flexibility on average (2.3 equations), with Spanish students having lower flexibility on average (1.4 equations). Middle school students had the lowest level of flexibility (0.5 equations), with advanced middle school students with slightly higher flexibility (1.2 equations) and advanced

mathematics track high school students having the greatest flexibility (3.6 equations). Swedish and Finnish advanced track high school students had the highest flexibility (each exhibiting flexibility on 4.3 equations, on average—or 36% of the equations on the assessment).

Note that this degree of procedural flexibility is quite consistent with prior studies. As noted above, in a study of experts' and undergraduates' performance on linear equations similar to those used in the present study, Lewis (1981) found that undergraduates chose to use a situationally appropriate strategy on only 12% of equations. Similarly, and also on similar linear equations, Star and Seifert (2006) found that middle school students used situationally appropriate strategies on only 2.5 out of 19 equations (or 13.2%). In this light, Swedish and Finnish advanced track high school students' levels of flexibility are unexpectedly high.

Potential flexibility

The results for students' potential flexibility largely mirror the procedural flexibility results (see Tables 2, 3). As a reminder, a student was deemed to have potential flexibility on a problem if the situationally appropriate strategy was provided but either (a) this strategy was not identified as the best (PF-AB), or (b) the standard algorithm was not used (PF-BC). On the whole, across all participants, on average students exhibited potential flexibility on 1.1 of the 12 problems on the assessment. High school students' potential flexibility was generally higher than that of middle school students, both for the whole sample as well as for Spain, Sweden, and Finland individually.

Looking more closely at the potential flexibility of students within each country, Spanish students present an interesting case in several ways. First, in the advanced mathematics classes, Spanish students had the highest potential flexibility both in middle school (0.9 equations) and in high school (1.1 equations). What appeared to prevent these students from being flexible was that they did not select the situationally appropriate strategy as better than the standard strategy.

Second, consider Spanish students' knowledge of the standard algorithm. Across all three countries, PF-BC (students who did not demonstrate knowledge of the standard algorithm) was rare. However, among Spanish advanced middle school students and advanced math track high school students, the average PF-BC score was 0—meaning that the advanced middle school and advanced track high school students consistently exhibited knowledge of the standard algorithm. By comparison, for Finland and Sweden, the PF-BC score was very small but positive, with weighted average of 0.04 and 0.10, respectively.

And third, Spanish students' procedural flexibility was also constrained by their choice of situationally appropriate strategies in the third phase of the test. Considering only those students who knew both the standard and situationally appropriate strategy (F + PF-AB) and looking at the proportion of these students who were not flexible (PF-AB), we find that in Spain, 33% of students who knew both types of strategies identified the standard strategy as “best” (and so were not deemed flexible), whereas the corresponding numbers for Finland and Sweden were 19% and 20%, respectively. A similar pattern holds when comparing middle school students to other middle school students and high school students to other high school students, except in the basic track high school group, where numbers are more similar across countries. We return to these points below in the discussion, specifically the differences between the strategies used and the procedural flexibility of Spanish students as compared to students from Finland and Sweden.

Spontaneous flexibility

Spontaneous flexibility, operationalized to be when students used a situationally appropriate strategy in the first attempt on a problem (in Phase 1), was very low across the sample (with an overall average of 0.5 problems out of 12) and virtually non-existent among Spanish students (the Spanish average across all students was 0.2; see Tables 2, 3). Spontaneous flexibility was highest for advanced track high school students, with Swedish advanced mathematics track students possessing surprisingly high levels (2.2 problems out of a possible 12). Across the entire sample, 162 students had at least one spontaneously flexible solution, with 36 students contributing 48% of the spontaneous flexibility cases. Spontaneous flexibility occurred on all tasks but was slightly more prevalent in tasks 1–3.

Analyses at the student level

As a separate analysis, we investigated characteristics of each student's performance on the test as a whole. Looking across all 12 equations, did students exhibit knowledge of the standard algorithm and/or a situationally appropriate strategy anywhere on the assessment? Did students exhibit flexibility and/or spontaneous flexibility on any of the equations on the assessment? This analysis reveals interesting differences between students in Finland, Sweden, and Spain.

With respect to the strategies that students used while equation solving, there is strong evidence that almost all participating students in each country had knowledge of the standard algorithm. As Table 4 indicates, 100% of Spanish high school students, more than 95% of Finnish and Swedish high school students, and more than 95%

Table 4 Percent of students from each country with at least one standard, situational, flexible, or spontaneously flexible solution

	Standard			Situationally Appropriate			Flexible			Spontaneous Flexible		
	MS	HS-B	HS-A	MS	HS-B	HS-A	MS	HS-B	HS-A	MS	HS-B	HS-A
Fin	61.3	95.1	100.0	40.9	42.6	93.2	12.9	19.7	81.6	3.2	1.6	37.9
Swe	96.6	–	98.9	50.6	–	93.9	23.0	–	81.8	3.5	–	58.0
Spa	98.2	100.0	100.0	22.6	60.0	68.1	12.8	48.6	42.6	1.8	11.4	2.1

Swedish HS-B (basic track high school) students are not included because of the small sample size

MS middle school; HS high school; HS-A advanced mathematics track high school; HS-B basic track high school; Fin Finland; Swe Sweden; Spa Spain

of Spanish and Swedish middle school students used the standard algorithm on at least one of the 12 equations on the assessment. Finnish middle school students were somewhat of an exception, with only 61% using the standard algorithm on at least one problem. With respect to the use of situationally appropriate strategies, as expected, high school students seemed more aware of situationally appropriate strategies than middle school students. Among Swedish students, over 90% of high school students used a situationally appropriate strategy on at least one equation, as compared to about 50% of middle school students. Situationally appropriate strategy use was comparable in Finland amongst advanced high school students (93% used a situationally appropriate strategy at least once), but this figure dropped to about 40% for middle school and basic track high school students. Spanish students were the least likely to use situationally appropriate strategies, with 60% of high school students and about 22% of middle school students using a situationally appropriate strategy at least once.

Swedish students emerged as particularly strong when looking at the results for flexibility and spontaneous flexibility. Over 80% of Swedish high school students in the sample exhibited flexibility and 58% of high school advanced track students exhibited spontaneous flexibility on at least one equation on the assessment. Note that the former is comparable to Finnish students but the latter is substantially higher. In addition, Spanish students' reliance on the standard algorithm is also highlighted in this analysis. Only 2% of Spanish advanced track high school students exhibited spontaneous flexibility on any problem on the assessment (using a situationally appropriate strategy on the first attempt in solving a problem), as compared with 38% and 58% of Finnish and Swedish advanced mathematics track students.

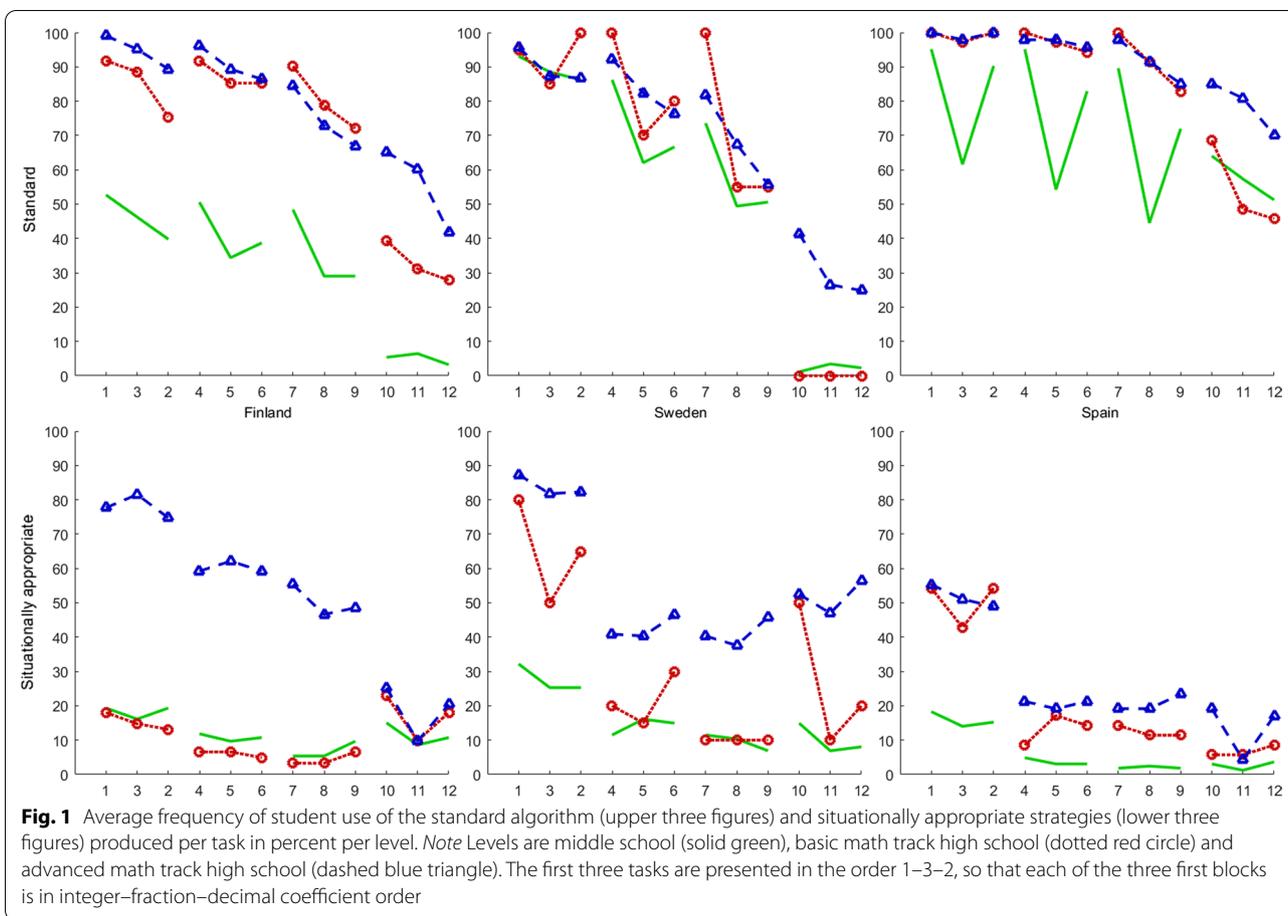
Problems attempted and strategies used

Although students in all three countries were allotted the same amount of time for the assessment, there were differences in how many problems students were able to attempt before time was called in each phase. In Phase 1 of the assessment, Spanish students on average attempted

10.2 of the 12 equations, as compared to 9.0 equations for Swedish students and 8.5 for Finnish students. Advanced mathematics track high school students in Spain were the quickest in progressing through the equations on the assessment in Phase 1, attempting an average of 11.0 of the 12 equations, while Finnish middle school students were the slowest, attempting on average only 6.6 equations. The proportion of participants who did not answer any of questions 10–12, correctly or incorrectly, was 35% for Finland, 33% for Sweden and 26% for Spain. The highest percentages were for Finnish and Swedish middle-school students (55–56%) and the lowest for Spanish advanced high-school students (13%). We believe many of these students ran out of time although some may simply not have known what to write for these problems. These results suggest that Spanish students who participated in the study were quicker equation solvers than those students in Finland and Sweden.

In Phase 2 of the assessment students were asked to provide additional strategies for solving each equation, including those equations that they did not attempt in the first phase. Over the two phases, Spanish students on average attempted more equations (an average of 10.2 problems) than did Finnish (9.3 problems) or Swedish (9.2 problems) students. Spanish advanced mathematics track high school students worked through the equations the quickest, attempting an average of 11.0 equations of the 12, while Finnish middle school students were again the slowest, attempting only an average of 7.3 equations.

Figure 1 provides additional information about the strategies that students used on the assessment. This figure shows the percentage of students in each country and at each level, where the standard algorithm (top row) or a situationally appropriate strategy (bottom row) was used for each equation number. For example, in the lower left pane, the point (9, 50) indicates that 50% of Finnish advanced math track high school students used a situationally appropriate strategy on Eq. 9 on the assessment, in either Phase 1 or Phase 2. Similarly, in the upper right pane, the point (3, 60) indicates that 60% of Spanish middle school students used the standard algorithm on Eq. 3 on the assessment. We note a general pattern on the



middle equation (Eqs. 3, 5, and 8) in each group of three, especially salient in Spanish (upper right pane) middle school students’ decreased use of the standard algorithm. The common factor here is that these are equations that involve fractional coefficients, which are well-known sources of difficulties for students (e.g., Siegler, 2017). The effect is slightly less pronounced for Finnish and Swedish students (upper left and middle pane), since they also had more difficulties with decimal coefficient problems (Eqs. 2, 6, and 9).

Figure 1 shows a number of interesting differences between the types of strategies used by Finnish, Swedish and Spanish students on the assessment. First, we see that Spanish students made extensive use of the standard algorithm (upper right pane) with very infrequent use of situationally appropriate strategies (lower right pane). Spanish students’ use of situationally appropriate strategies occurred primarily on Eqs. 1, 2, and 3 (i.e., the first equation type; see Table 1) of the assessment but very rarely on the other equation types.

Second, Fig. 1 also shows that Finnish basic track high school students’ (lower left pane, dotted red circle) use of

situationally appropriate strategies was almost identical to Finnish middle school students (lower left pane, solid green)—this similarity was less pronounced in Spain and Sweden. Third, we also note that the percentage of Swedish middle school students (upper middle pane, solid green) and basic track high school students (upper middle pane, dotted red circle) who use a standard algorithm drops dramatically on Problems 10, 11, and 12 of the assessment, perhaps indicating that Swedish students (other than those in the advanced high school track) may not have yet learned how to apply the standard algorithm to this equation type. Swedish students commonly used the standard algorithm on all other equations. Finally, there is a general and expected decline in the number of solutions towards the end of the test, but this effect is much more pronounced in the Finnish and Swedish samples.

In sum, with respect to equations attempted and strategies used, our results indicate considerable differences between Spanish, Finnish, and Swedish students—as well as within each country across ages and academic tracks. Spanish students across all ages and tracks used the

Table 5 Pearson correlations between accuracy and flexibility variables

	All countries			Finland			Sweden			Spain		
	All	MS	HS-A	All	MS	HS-A	All	MS	HS-A	All	MS	HS-A
<i>n</i>	791	344	331	257	93	103	288	87	181	246	164	47
<i>F</i>	0.60	0.32	0.51	0.70	0.63	0.46	0.70	0.41	0.66	0.38	0.33	0.22
SF	0.48	0.18	0.47	0.50	0.48	0.45	0.64	0.22	0.62	0.10	0.12	-0.04
F + PF	0.63	0.33	0.54	0.73	0.52	0.49	0.71	0.41	0.68	0.46	0.38	0.24

Bolded correlations are statistically significant at the $p < 0.0001$ level. Correlations above 0.5 are in bold italics

There were too few students in the high school basic tracks (HS-B) to calculate reliable correlations, but these students are included in the “All” columns. Note that the averages over countries are not weighted by the number of students in the different groups

MS middle school; HS-A advanced mathematics track high school; F flexibility; SF spontaneous flexibility; PF potential flexibility

standard algorithm more, and situationally appropriate strategies less, than either Finnish or Swedish students. In addition, Spanish students worked the fastest through the assessment, attempting the most equations.

Accuracy

Accuracy in solving the 12 linear equations on the assessment varied considerably. Across all groups of students in all three countries, on average half of equations were correctly solved in Phase 1 (average 5.4, see Table 2). As expected, high school students were more accurate, on average, than middle school students. Spanish high school students in the advanced mathematics track were the most accurate (averaging 8.7 equations correctly solved out of 12), followed by Finnish advanced math track high school students (7.9 equations solved correctly) and Swedish advanced math track high school students (7.5 solved correctly). Among the younger students, Spanish middle school students were again the most accurate, followed by Sweden and Finland. In fact, Spanish advanced middle school students (8.5 equations solved correctly) were even more accurate equation solvers than Finnish or Swedish advanced track high school students.

Noteworthy patterns were also seen when comparing the accuracies of middle and high school students from within the same country. In Finland, there was a large jump in accuracy between advanced middle school and advanced track high school students, with an average increase of 3.1 more equations solved correctly. There were more moderate jumps between middle school and basic math track high school students (in the range 1.8–2.3 for all countries). Interestingly, in Spain, there was virtually no change between advanced middle school students (8.5 problems solved correctly) and advanced track high school students (8.7 problems solved correctly).

In sum, the general findings from the analysis of students’ accuracy confirms the expected relationships between accuracy and grade level (middle school or high

school) as well as accuracy and academic track (advanced or basic) within each country. High school students were more accurate equation solvers than middle school students and advanced track high school students were more accurate than basic math track high school students. Results also show that Spanish students were the most accurate equation solvers, and that their ability to solve problems accurately was quite robust even among Spanish middle school students, with minimal average change as students moved into high school.

Relationships between procedural flexibility and accuracy

Finally, we sought to better understand the relationship between procedural flexibility and accuracy. We performed correlational analyses between accuracy scores and all variables related to procedural flexibility (see Tables 5, 6). Statistically significant correlations above 0.5 are generally considered to be strong (Cohen, 1988; Hemphill, 2003). However, note that we have elected to use the very conservative $p < 0.0001$ significance level, given that we perform 36 correlation calculations. Even with the most conservative (Bonferroni) correction, the

Table 6 Confidence intervals (95%) of Pearson correlations between accuracy and flexibility variables

		<i>n</i>	<i>F</i>	SF	F + PF
All countries	All	791	[0.56, 0.65]	[0.43, 0.54]	[0.58, 0.67]
	MS	344	[0.62, 0.75]	[0.40, 0.59]	[0.66, 0.78]
	HS-A	331	[0.27, 0.48]	[−0.02, 0.23]	[0.36, 0.55]
Finland	All	257	[0.66, 0.77]	[0.57, 0.71]	[0.66, 0.77]
	MS	93	[0.23, 0.42]	[0.07, 0.28]	[0.23, 0.42]
	HS-A	103	[0.49, 0.74]	[0.30, 0.62]	[0.36, 0.66]
Sweden	All	288	[0.19, 0.46]	[−0.03, 0.27]	[0.24, 0.50]
	MS	87	[0.22, 0.57]	[0.01, 0.41]	[0.22, 0.57]
	HS-A	181	[0.43, 0.59]	[0.38, 0.55]	[0.45, 0.61]
Spain	All	246	[0.29, 0.60]	[0.28, 0.59]	[0.33, 0.62]
	MS	164	[−0.07, 0.48]	[−0.32, 0.25]	[−0.05, 0.49]
	HS-A	47	[0.57, 0.74]	[0.52, 0.70]	[0.59, 0.75]

familywise error rate is 0.0036, which is less than 0.005. In addition, note that an increase in flexibility (F) has an ambiguous effect on potential flexibility (PF): A non-flexible student could become flexible (leaving the PF rate unchanged), or a student with PF could leave PF for F, thus reducing PF. For this reason, Tables 5, 6 includes the row F+PF which measures flexibility for potentially or “fully” flexible students. We use the Fisher *r*-to-*z* transformation to obtain an estimate of the significance of the differences between correlations.

On the whole, procedural flexibility and accuracy were strongly and positively correlated, $r = 0.60$. At the country level (and combining middle and high school students), procedural flexibility and accuracy were strongly correlated both in Finland ($r = 0.696$) and Sweden ($r = 0.704$); in Spain, this correlation was significant but lower ($r = 0.38$) than in the other two countries ($z > 5.20$; $p < 0.0001$). Among middle school students, only in Finland were procedural flexibility and accuracy strongly correlated ($r = 0.63$); in Spain and Sweden, the correlation was significant but of only moderate strength. Among high school students, only in Sweden were procedural flexibility and accuracy strongly correlated ($r = 0.66$); in Finland, this correlation was significant but lower ($z = 2.36$; $p < 0.02$). There was not a significant association between procedural flexibility and accuracy among Spanish high school students. A similar but weaker pattern emerges for spontaneous flexibility.

We interpret these results to indicate that flexibility and accuracy are generally related. The design of this study does not allow us to determine causality. However, regardless, it does appear that (a) greater accuracy is associated with greater procedural flexibility; (b) high accuracy is associated with the complete range of procedural flexibility, both high and low; and (c) without some ability to accurately solve equations, it is unlikely that a student will be flexible. However, our results also point to a substantial degree of variation—both within and between countries, as well as in terms of age and academic track.

Discussion

The goals of this study were to investigate procedural flexibility in linear equation solving among a convenience sample of middle and high school students in Spain, Finland, and Sweden. Despite its identification as an increasingly important mathematics learning outcome, the literature on flexibility and its development is relatively sparse. Lacking are international studies of students' procedural flexibility, particularly those that shed light on relationships that may exist between procedural flexibility, solving accuracy, and age/grade level. With the goal of addressing these gaps, our research questions in

the present study explored whether accuracy and procedural flexibility in linear equation solving varied by age, by academic track, and by country (RQ1 and RQ2), and the extent to which procedural flexibility and accuracy were related and how this relationship varied by age, by academic track, and by country (RQ3).

Our results indicate the following. First, we generally found expected relationships between students' age (middle school vs. high school) and both procedural flexibility and accuracy. High school students in all three countries were generally more accurate and more flexible solvers than middle school students. Similarly, when looking only at advanced mathematics track students, we found that high school students were generally more accurate and flexible solvers than advanced math middle school students.

Second, and consistent with prior research, we found that flexibility levels were modest among all students, with spontaneous flexibility quite low. Across all three countries and age groups, students exhibited flexibility on about 1.8 (about 15%) of the 12 problems, were potentially flexible on an additional 0.6 problems, and were spontaneously flexible on only 0.5 problems. Swedish and Finnish advanced track high school students had the highest flexibility, with each exhibiting flexibility on 36% (or about 4.3) of the equations on the assessment.

Third, our results indicate that knowledge and use of the standard algorithm for solving linear equations are widespread across all three countries. It appears that this standard algorithm is still the dominant method that is taught for solving equations, at least for the students in our sample from Spain, Sweden, and Finland. Relatedly, it also seems that the use of situationally appropriate methods for solving equations is present to some extent (particularly in Sweden and Finland) but is far less prevalent than the standard algorithm. In addition, fourth, we found that procedural flexibility and accuracy were positively and strongly correlated, where (in general) students with higher flexibility also had higher accuracy.

Fifth, our results also indicate interesting and important difference between Spanish, Swedish, and Finnish students. Spanish students were the most accurate solvers, and they achieved their high accuracy scores largely by implementing the standard algorithm. Spanish students' ability to accurately use the standard algorithm appeared to be already highly developed in middle school. Relatedly, Spanish students exhibited lower procedural flexibility and almost no spontaneous flexibility (when a student used a situationally appropriate strategy in the first solution attempt), indicating very infrequent use of situationally appropriate strategies. In addition, we found that the generally positive association between procedural flexibility and accuracy was weakest for Spanish

students, for whom there was no significant relationship between flexibility and accuracy. Furthermore, our results indicate that Finnish and Swedish students' equation solving profiles—including procedural flexibility and accuracy—were quite similar. Students in both countries had moderate flexibility (particularly for advanced track mathematics students in high school), some spontaneous flexibility, and reasonable accuracy. Swedish high school students in the basic track were much more flexible than their Finnish peers; however, the sample was small, so this conclusion is quite tentative.

It is perhaps not surprising that the results from Finnish and Swedish students are somewhat similar. Finland and Sweden have many cultural similarities, and perhaps one could presume that their educational systems with regard to mathematics instruction are also quite similar. Both countries have adopted similarly structured inclusive compulsory basic education; textbooks used by schools in each country have some basic similarities; and in both countries teachers are given considerable autonomy in terms of the choice of and implementation of curriculum (Hemmi & Krzywacki, 2014; Pehkonen et al., 2018). Yet at the same time, there are important differences between Finland and Sweden that one might presume would impact mathematics instruction in middle and high school, particularly procedural flexibility. For example, a comparative study about teacher education in Finland and Sweden (Hemmi & Ryve, 2015; see also Hemmi & Krzywacki, 2014) found differences in how mathematics teachers are educated. In particular, it was found that Swedish mathematics teacher educators prioritized the importance of individual interaction, building on students' ideas, and deriving mathematics from everyday situations, while in Finland, clarity, pedagogical routines, homework, and specific goals for every lesson were more highly valued. Another comparative study between Sweden and Finland found that in Swedish classrooms it is common to see students working through their textbooks independently at their own individualized pace, whereas Finnish teachers favor whole-class instruction, where all students are engaged with the same mathematical tasks (Pehkonen et al., 2018). Furthermore, a comparative study of Swedish and Finnish high school mathematics textbooks (Bergwall & Hemmi, 2017) revealed that Finnish textbooks offered more opportunities for learning proof, while Swedish materials featured a higher variation in the nature and types of proof-related reasoning. One might hypothesize that the more individualized approach that is more typical in Sweden would offer fewer opportunities to compare different strategies and thus to cultivate procedural flexibility, yet the present results appear to indicate the opposite—that on average Swedish students have slightly more procedural flexibility

than Finnish students. Understanding the relationships between instructional environments in mathematics classrooms in Finland and Sweden and these students' procedural flexibility is an interesting area for continued study.

With respect to the case of Spain, why might Spanish students' flexibility and accuracy in equation solving be so different than similarly aged students from Sweden and Finland? A main finding from the present study was that Spanish students (even beginning in middle school) were faster and more accurate equation solvers who tended to exclusively rely upon the standard algorithm. Unlike in Sweden and Finland, secondary mathematics teaching in Spain appears to be almost completely centered on routine application of methods to get the correct answer as fast as possible (González-Astudillo & Sierra-Vázquez, 2004). Spanish students are also used to working under time pressure to solve mathematics problems, often in the form of frequent timed written tests as well as standardized multiple-choice assessments (Rico, 1993). Spanish students' relatively greater facility with efficient use of standard algorithms has not previously been reported in international comparative studies. Anecdotal reports on Spanish equation solving (e.g., Joglar et al., 2018; Ruiz & Bosch, 2007) suggest that teachers place great emphasis on knowledge of and efficient use of standard algorithms and that Spanish students receive a lot of training and practice in how to quickly execute the standard algorithm. In addition, these results may be particularly influenced by the organization and design of the curriculum and the use of didactic materials in mathematics classrooms, especially textbooks. Despite some changes in the official mathematics curriculum, there is a great deal of inertia and resistance to change in Spanish instruction. For example, the algorithm for computing square roots, which has not appeared in the Spanish curriculum since 1997, is still present in many Spanish secondary mathematics classrooms. It is also the case that teachers, authors, and publishers understand textbooks as encyclopedic manuals, with lists of axioms, definitions, theorems, proofs, and routine exercises. As a result, most Spanish mathematics textbooks generally have a very methodical approach that prioritizes the use of traditional algorithms (López Beltrán et al., 2020).

In fact, it may even be the case that Spanish teachers and students believe that standard approaches are more highly valued than the situationally appropriate approaches used in the present study. Spanish students might perceive that the "best" strategy is the one that reduces the likelihood of making mistakes or the one that the teachers and textbooks use more frequently, which may make the standard algorithm for solving linear equations the best default strategy for them. In support

of this latter point, our results indicate that Spanish students had almost twice as high a chance of choosing the standard over the situationally appropriate strategy as compared to the other two countries (33% vs. 19%). Relatively high PF-AB scores (e.g., on 1.1 of the 12 problems for Spanish advanced high school students, the highest PF-AB score for students in any country or age group) suggests that Spanish students, when asked to indicate whether the standard algorithm or a situationally appropriate strategy was best, chose the standard strategy.

The idea that a standard algorithm could be considered as the best approach to a problem, perhaps even better than what we refer to here as a situationally appropriate approach, has a rational basis. After all, the standard algorithm is applicable to a wide range of equation types, while a particular situationally appropriate strategy is applicable to a much narrower range of equations. As a result, it is generally possible to immediately start using a standard algorithm on a given equation, avoiding the need to examine a problem's structural features to determine whether one of several possible situationally appropriate approaches could be used. Thus, if a standard algorithm has become well-practiced, a student may be able to quickly, effortlessly, and accurately implement it—and thus, it may indeed be the best approach to that problem, especially in a test-like context like the Tri-Phase Flexibility Assessment used here. This type of solving behavior—a strong preference for the standard algorithm, even amongst students who exhibited knowledge of both standard and situationally appropriate approaches—was relatively uncommon amongst Finnish and Swedish students in this study, as well as among Chinese students in a prior study using the same instrument (Xu et al., 2017). This suggests that it may be promising to further explore this issue. In particular, what criteria do students use in determining which strategies are the best? What role does prior mathematical experience play in this determination? Previous studies have found that mathematics experts showed relatively consistent criteria for identifying the most elegant or situationally appropriate strategy as the best, while students held more diverse criteria; furthermore, some students also believed the standard strategy was better (as did many of the Spanish students in the present study) (Star & Madnani, 2004; Star & Newton, 2009). Researchers have suggested that such knowledge of situationally appropriate strategies might be related to both the sociocultural context of the learner and the learners' prior knowledge (Hatano & Inagaki, 1984; Newton et al., 2019; Verschaffel et al., 2009). Thus, it would be interesting to further investigate how students evaluate which strategy is the “best” and how cross-cultural factors and prior knowledge might affect students' evaluation criteria.

Finally, it is also the case that differences in teacher preparation programs between the three countries could be a possible explanation for the presumed variation in teaching practices that may underlie some of our findings. Recent studies (Muñiz-Rodríguez et al., 2016), based on the TEDS-M study, show that secondary mathematics teacher preparation programs in Spain may be relatively ineffective, at least as compared to those in other countries, including Finland. Designing interventions to address the problem of how to promote students' procedural flexibility as part of professional development programs in Spain might be an interesting direction for further work.

Limitations

The following limitations suggest caution in the interpretation of the results of this study. First, in this descriptive study, we relied on a relatively small convenience sample of students in each of the three countries that were examined. We also were unable to obtain demographic information about participating students including gender, race, and socio-economic status. Future studies in this area should consider utilizing a larger and/or a representative sample that includes a broad range of student demographic indicators. The relationship between flexibility and these demographic factors is largely unexplored (cf. Star et al., 2015). However, given prior work that more generally establishes links between mathematics achievement and students' race, gender, and socio-economic status (e.g., McGraw et al., 2006; Reyes & Stanic, 1988), it is reasonable to assume that such a relationship would also exist for flexibility and is worthy of further investigation. Second, the design of the study was cross-sectional, which clearly limits our ability to make any substantive claims about the development of procedural flexibility between middle and high school; the present results are merely suggestive. Third, our focus here was on procedural flexibility in the mathematical domain of linear equation solving; it would be instructive to continue to explore flexibility and its development in multiple countries in other or multiple mathematical domains. Fourth, the assessment that was used to measure flexibility was a timed test. As a result, some students may not have had sufficient time to list all alternative strategies that they knew for a given problem (see also Hästö & Palkki, 2019, for an in-depth analysis of student speed in the Finnish part of the data set). Furthermore, timed tests have been linked by some scholars to increased math anxiety, which in turn can negatively impact students' performance (Boaler, 2014). Thus it may be the case that students would have exhibited greater flexibility had the assessment been administered in an untimed manner. Fifth, because of the different educational system being

examined, it was not always possible to compare identical samples; for example, the middle school students in the present study were in 8th grade in Spain and Finland but in 9th grade in Sweden. Such differences in age across countries could explain some of the differences in flexibility that were found here. Finally, the distribution of students in each country at each level varied considerably, which could have impacted results. For example, in Spain only 47 of the 246 students (19%) were in the advanced mathematics track in high school, while in Finland and Sweden, these proportions were 40% (103 of 258) and 63% (181 of 288), respectively.

Conclusions

Although it is increasingly recognized as an important construct of interest, there are many aspects of mathematical flexibility that are not well-understood. Particularly lacking in the literature on flexibility are studies that explore similarities and differences in students' repertoire of strategies for solving algebra problems across countries with different educational systems and curricula. Although its use of a cross-sectional design and a convenience sample of students limits the generalizability of its results, this study yielded important insights about flexibility, including widespread use of a standard algorithm among participating students in all three countries that were investigated, substantial variation in the use of standard as well as situationally appropriate approaches to linear equation solving, and positive correlations between the accuracy of students' strategies and their flexible use of these strategies. Future research can continue to explore the extent that within- and between-country differences in flexibility can be linked to differences in countries' educational systems, teaching practices, and/or cultural norms around mathematics teaching and learning.

Abbreviations

NCTM: National Council of Teachers of Mathematics; PF: Potential flexibility; SF: Spontaneous flexibility; F: Flexibility; MS: Middle school; HS-A: Advanced mathematics track high school; HS-B: Basic track high school; Fin: Finland; Swe: Sweden; Spa: Spain.

Acknowledgements

Not applicable.

Authors' contributions

Data collection and analysis for the Spanish data was conducted by NP and MA. Data collection and analysis for the Swedish data was conducted by JP. Data collection and analysis for the Finnish data was conducted by DT, PH, and RP. Additional analyses were conducted by PH, DT, RJ, RL, and LL. JS was primarily responsible for the design of the study and instruments as well as the writing of the paper. All authors read and approved the final manuscript.

Funding

Not applicable.

Availability of data and materials

The data sets used and/or analyzed during the current study are available from the corresponding author on reasonable request.

Declarations

Competing interests

The authors declare that they have no competing interests.

Author details

¹Harvard Graduate School of Education, 442 Gutman Library, Cambridge, MA 02138, USA. ²University of Oulu, Oulu, Finland. ³Universidad Complutense de Madrid, Madrid, Spain. ⁴Universidad Rey Juan Carlos, Madrid, Spain. ⁵Chalmers University of Technology and the University of Gothenburg, Gothenburg, Sweden. ⁶Beijing Normal University, Beijing, China.

Received: 6 May 2021 Accepted: 25 December 2021

Published online: 10 January 2022

References

- Bergwall, A., & Hemmi, K. (2017). The state of proof in Finnish and Swedish mathematics textbooks—Capturing differences in approaches to upper-secondary integral calculus. *Mathematical Thinking and Learning*, *19*(1), 1–18. <https://doi.org/10.1080/10986065.2017.1258615>
- Blöte, A. W., Van der Burg, E., & Klein, A. S. (2001). Students' flexibility in solving two-digit addition and subtraction problems: Instruction effects. *Journal of Educational Psychology*, *93*(3), 627. <https://doi.org/10.1037/0022-0663.93.3.627>
- Boaler, J. (2014). Research suggests that timed tests cause math anxiety. *Teaching Children Mathematics*, *20*(8), 469–474.
- Boli, J. (2014). *New citizens for a new society: The institutional origins of mass schooling in Sweden*. Elsevier.
- Booth, J. L., Barbieri, C., Eyer, F., & Paré-Blagoev, E. J. (2014). Persistent and pernicious errors in algebraic problem solving. *The Journal of Problem Solving*, *7*(1), 3. <https://doi.org/10.7771/1932-6246.1161>
- Buchbinder, O., Chazan, D., & Fleming, E. (2015). Insights into the school mathematics tradition from solving linear equations. *For the Learning of Mathematics*, *35*, 2–8.
- Bush, S. B., & Karp, K. S. (2013). Prerequisite algebra skills and associated misconceptions of middle grade students: A review. *The Journal of Mathematical Behavior*, *32*(3), 613–632. <https://doi.org/10.1016/j.jmathb.2013.07.002>
- Cai, J. (2000). Mathematical thinking involved in US and Chinese students' solving of process-constrained and process-open problems. *Mathematical Thinking and Learning*, *2*(4), 309–340. https://doi.org/10.1207/S15327833MTL0204_4
- Cai, J. (2004). Why do US 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>
- Cai, J., & Hwang, S. (2002). Generalized and generative thinking in US and Chinese students' mathematical problem solving and problem posing. *The Journal of Mathematical Behavior*, *21*(4), 401–421. [https://doi.org/10.1016/S0732-3123\(02\)00142-6](https://doi.org/10.1016/S0732-3123(02)00142-6)
- Carr, M., & Taasobshirazi, G. (2017). Is strategy variability advantageous? It depends on grade and type of strategy. *Learning and Individual Differences*, *54*, 102–108. <https://doi.org/10.1016/j.lindif.2017.01.015>
- Chen, C., & Stevenson, H. W. (1995). Motivation and mathematics achievement: A comparative study of Asian-American, Caucasian-American, and East Asian high school students. *Child Development*, *66*(4), 1215–1234. <https://doi.org/10.1111/j.1467-8624.1995.tb00932.x>
- Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Hillsdale.
- De Bock, D., Van Dooren, W., Janssens, D., & Verschaffel, L. (2002). Improper use of linear reasoning: An in-depth study of the nature and the irresistibility of secondary school students' errors. *Educational Studies in Mathematics*, *50*(3), 311–334. <https://doi.org/10.1023/A:1021205413749>

- De Bock, D., Van Dooren, W., Janssens, D., & Verschaffel, L. (2007). *The illusion of linearity: From analysis to improvement*. Springer.
- Dover, A., & Shore, B. M. (1991). Giftedness and flexibility on a mathematical set-breaking task. *Gifted Child Quarterly*, *35*(2), 99–124. <https://doi.org/10.1177/001698629103500209>
- Dowker, A. (1992). Computational estimation strategies of professional mathematicians. *Journal for Research in Mathematics Education*, *23*, 45–55. <https://doi.org/10.2307/749163>
- Fernández, C., Llinares, S., Van Dooren, W., De Bock, D., & Verschaffel, L. (2012). The development of students' use of additive and proportional methods along primary and secondary school. *European Journal of Psychology of Education*, *27*(3), 421–438. <https://doi.org/10.1007/s10212-011-0087-0>
- Flavell, J. H., & Wohlwill, J. F. (1969). Formal and functional aspects of cognitive development. In D. Elkind & J. H. Flavell (Eds.), *Studies in cognitive development: Essays in honor of Jean Piaget* (pp. 67–120). Oxford University Press.
- 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*, *22*(3), 705. in Spanish.
- Gorgorió, N., Albarracín, L., Laine, A., & Llinares, S. (2018). Alicante-Helsinki-Barcelona: Students' mathematical background and requirements to enter primary teaching degree. In E. Bergqvist, M. Österholm, C. Granberg, & L. Sumpter (Eds.), *Proceedings of the 42nd Conference of the International Group for the Psychology of Mathematics Education* (Vol. 5, p. 237). Umeå, Sweden: PME.
- Hardy, G. H. (1940). *A Mathematician's apology*. Cambridge University Press.
- Hästö, P., & Palkki, R. (2019). Finnish students' flexibility and its relation to speed and accuracy in equation solving. *Nordic Studies in Mathematics Education*, *24*(3–4), 43–58.
- Hästö, P., Palkki, R., Tuomela, D., & Star, J. R. (2019). Relationship between mathematical flexibility and success in national examinations. *European Journal of Science and Mathematics Education*, *7*(1), 1–13.
- Hatano, G., & Inagaki, K. (1984). Two courses of expertise. *Research and Clinical Center for Child Development Annual Report*, *6*, 27–36.
- Hemmi, K., & Krzywacki, H. (2014). Crossing the boundaries: Swedish teachers' interplay with Finnish curriculum materials. In *Conference on Mathematics Textbook Research and Development (ICMT-2014)* (p. 263).
- Hemmi, K., & Ryve, A. (2015). Effective mathematics teaching in Finnish and Swedish teacher education discourses. *Journal of Mathematics Teacher Education*, *18*(6), 501–521. <https://doi.org/10.1007/s10857-014-9293-4>
- Hemphill, J. F. (2003). Interpreting the magnitudes of correlation coefficients. *American Psychologist*, *58*, 78–79. <https://doi.org/10.1037/0003-066X.58.1.78>
- Huntley, M. A., Marcus, R., Kahan, J., & Miller, J. L. (2007). Investigating high-school students' reasoning strategies when they solve linear equations. *The Journal of Mathematical Behavior*, *26*(2), 115–139. <https://doi.org/10.1016/j.jmathb.2007.05.005>
- 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>
- Joglar, N., Abánades, M., & Star, J. R. (2018). April. Flexibilidad matemática y resolución de ecuaciones lineales. *Uno: Revista De Didáctica De Las Matemáticas*, *80*, 51–57.
- Lemaire, P., & Siegler, R. S. (1995). Four aspects of strategic change: Contributions to children's learning of multiplication. *Journal of Experimental Psychology: General*, *124*(1), 83–97. <https://doi.org/10.1037/0096-3445.124.1.83>
- Levav-Waynberg, A., & Leikin, R. (2012). The role of multiple solution tasks in developing knowledge and creativity in geometry. *The Journal of Mathematical Behavior*, *31*(1), 73–90. <https://doi.org/10.1016/j.jmathb.2011.11.001>
- Lewis, C. (1981). Skill in algebra. In J. R. Anderson (Ed.), *Cognitive skills and their acquisition* (pp. 85–110). Lawrence Erlbaum Associates.
- Li, X., Jiang, R., & Qian, Y. (2014). 5–8 Graders' overuse of proportionality on missing-value problems. *Journal of Mathematics Education*, *23*(6), 73–77.
- Liu, R., Wang, J., Star, J. R., Zhen, R., Jiang, R., & Fu, X. (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 Beltrán, M., Albarracín Gordo, L., Ferrando Palomares, I., Montejo-Gómez, J., Ramos Alonso, P., Serradó Bayés, A., Thibaut Tadeo, E., & Mallavibarrena, R. (2020). La educación matemática en las enseñanzas obligatorias y el bachillerato. In D. Martín de Diego (Ed.), *Libro Blanco de las Matemáticas* (pp. 1–94). Real Sociedad Matemática Española y Fundación Ramón Areces. <https://www.fundacionareces.es/fundacionareces/es/publicaciones/libro-blanco-de-las-matematicas.html>
- Lynch, K., & Star, J. R. (2014). Views of struggling students on instruction incorporating multiple strategies in Algebra I: An exploratory study. *Journal for Research in Mathematics Education*, *45*(1), 6–18. <https://doi.org/10.5951/jresmetheduc.45.1.0006>
- Maciejewski, W. (2020). Between confidence and procedural flexibility in calculus. *International Journal of Mathematical Education in Science and Technology*. <https://doi.org/10.1080/0020739X.2020.1840639>
- Maciejewski, W., & Star, J. (2016). Developing flexible procedural knowledge in undergraduate calculus. *Research in Mathematics Education*, *18*(3), 299–316. <https://doi.org/10.1080/14794802.2016.1148626>
- Maciejewski, W., & Star, J. R. (2019). Justifications for choices made in procedures. *Educational Studies in Mathematics*, *101*, 325–340. <https://doi.org/10.1007/s10649-019-09886-7>
- Martin, M. O., Mullis, I. V., Foy, P., & Olson, J. (2008). *TIMSS 2007 International Mathematics Report*. Chestnut Hill: TIMSS & PIRLS International Study Center, Boston College.
- Martin, T., Rayne, K., Kemp, N. J., Hart, J., & Diller, K. R. (2005). Teaching for adaptive expertise in biomedical engineering ethics. *Science and Engineering Ethics*, *11*(2), 257–276.
- McGraw, R., Lubienski, S. T., & Strutchens, M. E. (2006). A closer look at gender in NAEP mathematics achievement and affect data: Intersections with achievement, race/ethnicity, and socioeconomic status. *Journal for Research in Mathematics Education*, *37*(2), 129–150.
- McMullen, J., Brezovszky, B., Hannula-Sormunen, M. M., Veermans, K., Rodríguez-Aflecht, G., Pongsakdi, N., & Lehtinen, E. (2017). Adaptive number knowledge and its relation to arithmetic and pre-algebra knowledge. *Learning and Instruction*, *49*, 178–187. <https://doi.org/10.1016/j.learninstruc.2017.02.001>
- Mullis, I. V., Martin, M. O., Foy, P., & Arora, A. (2012). *TIMSS 2011 International Results in Mathematics*. Herengracht 487, Amsterdam, 1017 BT, The Netherlands: ERIC.
- Muñiz-Rodríguez, L., Alonso, P., Rodríguez-Muñiz, L. J., & Valcke, M. (2016). ¿Hay un vacío en la formación inicial del profesorado de matemáticas de Secundaria en España respecto a otros países? [Is there a gap in initial secondary mathematics teacher education in Spain compared to other countries?]. *Revista de educación*, *372*, 106–132. <https://doi.org/10.4438/1988-592X-RE-2015-372-317>
- Mustonen, V., & Hakkarainen, K. (2015). Tracing two apprentices' trajectories toward adaptive professional expertise in fingerprint examination. *Vocations and Learning*, *8*(2), 185–211.
- National Council of Teachers of Mathematics. (2014). *Procedural fluency in mathematics: A position of the National Council of Teachers of Mathematics*. Retrieved from <https://www.nctm.org/Standards-and-Positions/Position-Statements/Procedural-Fluency-in-Mathematics/>.
- National Research Council. (2001). Adding it up: Helping children learn mathematics. In J. Kilpatrick, J. Swafford, & B. Findell (Eds.), *Mathematics Learning Study Committee, Center for Education, Division of Behavioral and Social Sciences and Education*. National Academy Press.
- 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*. <https://doi.org/10.1080/00220973.2019.1586629>
- Newton, K., Star, J. R., & Lynch, K. (2010). Understanding the development of flexibility in struggling algebra students. *Mathematical Thinking and Learning*, *12*(4), 282–305. <https://doi.org/10.1080/10986065.2010.482150>
- Pehkonen, L., Hemmi, K., Krzywacki, H., & Laine, A. (2018). A cross-cultural study of teachers' relation to curriculum materials. In *Nordic Research in Mathematics Education Papers of NORMA 17 The Eighth Nordic Conference on Mathematics Education Stockholm, May 30-June 2, 2017*. Svensk förening för matematikdidaktisk forskning.
- Pirttimaa, M., Husu, J., & Metsärinne, M. (2017). Uncovering procedural knowledge in craft, design, and technology education: A case of hands-on

- activities in electronics. *International Journal of Technology and Design Education*, 27(2), 215–231.
- Reyes, L. H., & Stanic, G. M. (1988). Race, sex, socioeconomic status, and mathematics. *Journal for Research in Mathematics Education*, 19(1), 26–43.
- Rico, L. (1993). Mathematics Assessment in the Spanish Educational System. In M. Niss (Ed.) *Cases of Assessment in Mathematics Education*. New ICMI Study Series, vol 1. Springer. https://doi.org/10.1007/978-94-017-0980-4_2
- 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>
- Rittle-Johnson, B., & Star, J. R. (2009). Compared with what? The effects of different comparisons on conceptual knowledge and procedural flexibility for equation solving. *Journal of Educational Psychology*, 101(3), 529–544. <https://doi.org/10.1037/a0014224>
- Rittle-Johnson, B., Star, J. R., & Durkin, K. (2009). The importance of prior knowledge when comparing examples: Impact on conceptual and procedural knowledge of equation solving. *Journal of Educational Psychology*, 101(4), 836–852. <https://doi.org/10.1037/a0016026>
- Rittle-Johnson, B., Star, J. R., & Durkin, K. (2012). Developing procedural flexibility: When should multiple procedures be introduced? *British Journal of Educational Psychology*, 82, 436–455.
- Ruiz, N., & Bosch, J. (2007). La educación matemática en España. *Práxis Educativa*, 2(2), 151–160.
- Sahlberg, P. (2014). *Finnish Lessons 2.0: What Can the World Learn From Educational Change in Finland?*. Teachers College Press.
- 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, 981.
- Siegler, R. S., & Lortie-Forgues, H. (2017). Hard lessons: Why rational number arithmetic is so difficult for so many people. *Current Directions in Psychological Science*, 26(4), 346–351. <https://doi.org/10.1177/0963721417700129>
- Star, J., & Madnani, J. (2004). *Which way is best? Students' conceptions of optimal strategies for solving equations*. Paper presented at the Proceedings of the twenty-sixth annual meeting of the North American chapter of the International Group for the Psychology of Mathematics Education.
- Star, J. R. (2005). Reconceptualizing procedural knowledge. *Journal for Research in Mathematics Education*, 36(5), 404–411. <https://doi.org/10.2307/30034943>
- Star, J., & 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., Newton, K., Pollack, C., Kokka, K., Rittle-Johnson, B., & Durkin, K. (2015). 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., & Rittle-Johnson, B. (2008). Flexibility in problem solving: The case of equation solving. *Learning and Instruction*, 18, 565–579. <https://doi.org/10.1016/j.learninstruc.2007.09.018>
- Star, J. R., & Rittle-Johnson, B. (2009). It pays to compare: An experimental study on computational estimation. *Journal of Experimental Child Psychology*, 102, 408–426. <https://doi.org/10.1016/j.jecp.2008.11.004>
- Star, J. R., & Seifert, C. (2006). The development of flexibility in equation solving. *Contemporary Educational Psychology*, 31, 280–300. <https://doi.org/10.1016/j.cedpsych.2005.08.001>
- Torbeys, J., De Smedt, B., Ghesquière, P., & Verschaffel, L. (2009). Jump or compensate? Strategy flexibility in the number domain up to 100. *ZDM Mathematics Education*, 41(5), 581–590. <https://doi.org/10.1007/s11858-009-0187-3>
- Uttal, D. H., Miller, D. I., & Newcombe, N. S. (2013). Exploring and enhancing spatial thinking: Links to achievement in science, technology, engineering, and mathematics? *Current Directions in Psychological Science*, 22(5), 367–373.
- Verschaffel, L., Luwel, K., Torbeys, J., & Van Dooren, W. (2009). Conceptualizing, investigating, and enhancing adaptive expertise in elementary mathematics education. *European Journal of Psychology of Education*, 24, 335–359. <https://doi.org/10.1007/BF03174765>
- Xu, L., Liu, R., Star, J. R., Wang, J., Liu, Y., & Zhen, R. (2017). Measures of potential flexibility and practical flexibility in equation solving. *Frontiers in Psychology*, 8, 1–13. <https://doi.org/10.3389/fpsyg.2017.01368>

Publisher's Note

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

Submit your manuscript to a SpringerOpen[®] journal and benefit from:

- Convenient online submission
- Rigorous peer review
- Open access: articles freely available online
- High visibility within the field
- Retaining the copyright to your article

Submit your next manuscript at ► [springeropen.com](https://www.springeropen.com)