

University students' strategy flexibility in equation solving

Anne-Maria Ernvall-Hytönen^{a1}, Peter Hästö^a, Heidi Krzywacki^b and Sara Parikka^a

^a*Department of Mathematics and Statistics, University of Helsinki, Helsinki, Finland;*

^b*Department of Education, University of Helsinki, Helsinki, Finland*

This study focuses on students' strategy flexibility when solving linear and quadratic equations in an introductory university mathematics course. In this setting, we relate results from a tri-phase flexibility test with course achievement. We find that strategy flexibility is related to course achievement for above average students. The results show statistically significant differences in flexibility between degree programs and high-school mathematics backgrounds. In addition, we compare the flexibility of the university students with that of high school students from an earlier dataset and find a substantial differences in both strategy flexibility and equation solving accuracy.

Keywords: strategy flexibility, procedural flexibility, adaptability, equation solving, university mathematics, mathematics achievement

Introduction

Flexibility is an important aspect of mathematical competence that is explicitly recognized in many mathematics education policy documents (e.g., NCTM, 2023), albeit not in Finland (FNBE, 2014; FNBE, 2019), the context of this research. Nevertheless, Finnish pupils have shown similar levels of flexibility as pupils in other countries (Star et al., 2022; Xu et al., 2017) and flexibility is a popular topic for teachers' continued professional development courses in Finland.

Mathematical flexibility has been widely studied in Europe, North America and China (Hong et al., 2023; Verschaffel, 2024). Researchers have considered it from different perspectives, including flexible use of strategies and representations, and in various domains (Heinze et al., 2009; Hickendorff et al., 2022). Most research on strategy flexibility concerns primary and secondary grades (Hästö & Star, 2024) and Verschaffel (2024) points out that different insights may be gleaned from examining more advanced mathematical students. To address this gap, this paper concerns strategy flexibility of university students. To state our research questions, we begin with an overview of the relevant literature and a discussion on the definition of strategy flexibility.

Literature and theoretical framework

Star (2005) argued that procedural knowledge is not necessarily superficial and suggested that flexibility in the choice and application of solution strategies is an example of deep procedural knowledge. Much work has been done since then, as reported by Hong et al. (2023), Cox et al. (2024) and Verschaffel (2024) in their recent literature reviews. However, in many ways the theory of procedural flexibility is still in its early stages of development, and still features variability in terminology. Cox et al. (2024) state that, despite the importance of mathematical flexibility, a clear definition is missing. They identify four themes variously defining mathematical flexibility in the literature: adaptation, use of multiple strategies, efficient use of strategies, and

¹ Corresponding author: anne-maria.ernvall-hytonen@helsinki.fi, PL 68, Pietari Kalmin katu 5, 00014 Helsingin yliopisto, Finland

Author are listed in alphabetical order reflecting a substantial contribution by each.

identification of problem features. These pertain specifically to a form of mathematical flexibility known as strategy flexibility or procedural flexibility. In this paper, we focus on this kind of flexibility. In his recent literature review, Verschaffel (2024), following Star and Newton (2009), defined *strategy flexibility* by two criteria:

- (SF1) knowledge of multiple strategies, and
- (SF2) ability and tendency to select the most appropriate strategy for the problem at hand.

Assessing strategy appropriateness for criterion (SF2) is non-trivial (Verschaffel, 2024). The mathematical characteristics of the problem are naturally considered since using a strategy regardless of the problem's nature epitomizes inflexibility. However, some researchers, especially at the elementary school level, also consider characteristics of the person solving the problem and the sociocultural context in which the work is carried out (Hong et al. 2023; Hickendorff, 2022; Verschaffel, 2024). Note that not considering personal or social factors in the definition is not the same as claiming that they do not impact flexibility. Rather, considering only problem characteristics means that we investigate how good students are at choosing strategies that are mathematically suitable for the problem at hand. Personal or social factors may imply a low probability of choosing such strategies, corresponding to low flexibility. If these factors are included in the definition, it means that we instead strive to determine if the students strategy choice aligns with his or her proficiency with different strategies. If the student has strong routine skills with the mathematically basic strategy, it may then deemed strategic to chose it over a mathematically more appropriate strategy that the student or community is less familiar with. In sum, considering personal and social factors places more emphasis on metacognitive skills whereas not including them emphasizes mathematical competence. We refer to recent extensive literature reviews for further discussions about the definition of flexibility (Cox et al., 2024; Hong et al. 2023; Verschaffel, 2024).

We remark that the relevant problem characteristics for what constitutes flexibility will naturally differ between different mathematical domains. In this view, being "flexible" is not so much a trait as a skill (deep procedural knowledge) that manifests itself in solving certain tasks. It is quite possible, indeed expected, that one could show flexibility in a familiar domain but not in a an unfamiliar one. With the recent development of more efficient measures of strategy flexibility (Hong et al. 2023; Verschaffel, 2024), we expect future studies to address this issue directly. In this paper we take a small step in this direction by extending previous test from using only linear equations to include also quadratic equations. Since our primary interest is in flexibility as a mathematical competence, we assess appropriateness only based on problem characteristics. We therefore use a tri-phase method in this study, where appropriateness of a strategy is assessed based only on problem characteristics (e.g., Xu et al., 2017; Star et al., 2022). This method is described below in the section "Materials and test procedure".

Literature overview and research questions

Verschaffel (2024) identifies several areas where strategy flexibility has not been investigated or where results have been inconclusive. He concludes that little research directly addresses the relationship between strategy flexibility and mathematics achievement or mathematical ability and that it has yielded mixed findings (e.g., Hästö et al., 2019; Star et al., 2022; Torbeyns et al., 2017). In support of a positive relation, interventions on linear equation solving led to gains in both strategy flexibility and

problem-solving accuracy (Garcia Coppersmith & Star, 2022; Rittle-Johnson et al., 2009, 2012; Rittle-Johnson & Star, 2007, 2009; Star & Rittle-Johnson, 2008). In studies with the tri-phase method, this positive relationship was not found in Chinese middle-school students (Xu et al., 2017) and occurred to different extents in middle- and high-school students in Finland, Spain and Sweden (Star et al., 2022). Especially relevant for the current study is the finding by Bye et al. (2022) of no relation between US undergraduates' strategy flexibility in arithmetic equation solving measured through reaction times to simple missing number problems like $x - 2 = 7$ and their standardized mathematics achievement scores.

Previous research has also been inconsistent regarding the relationship between students' grade-level and flexibility: some studies find higher grade-level to be related to higher strategy flexibility (e.g., Star et al., 2022; Torbeyns et al., 2009) whereas others have found no connection between the two (e.g., Keleş & Yazgan, 2021; Van Der Auwera et al., 2022).

To the best of our knowledge, there are only six papers dealing with flexibility at the university level, so we briefly review them all here. Maciejewski & Star (2016) have shown that a short intervention can lead to strategy flexibility in simple differentiation tasks, whereas Maciejewski (2022) found a correlation between strategy flexibility in calculus and self-confidence. In linear algebra, Kontorovich (2020) suggested that flexibility is required to use appropriate theorems instead of falling back on Gaussian elimination as a standard algorithm. Similarly, Broley and Hardy (2022) found that students were mostly inflexible with a strong predilection for calculus routines over more efficient analysis tools. In line with Bye et al. (2022) mentioned above, Shaw et al. (2020) found no connection between undergraduate students' numbers of solutions for simple arithmetic problems (a proxy of flexibility) and affective variables, such as mathematics anxiety, or working memory.

The term spontaneous (or practical) flexibility has been used to describe specifically the tendency to select appropriate strategies in (SF2) of the definition of strategy flexibility. Research has found very low levels of spontaneity in the flexible use of strategies in algebra tasks, which made studying its impact unreliable (e.g. Hästö & Palkki, 2019; Hickendorff, 2018; Star et al., 2022; Xu et al., 2017). However, as university students are more mathematically advanced, we are also able to consider this aspect of flexibility.

Despite the prominence of linear equations in secondary school flexibility research (e.g., Huntley et al., 2007; Star et al., 2022; Star & Rittle-Johnson, 2008; Xu et al., 2017), strategy flexibility in algebra has not been considered at the university level. Furthermore, Star et al. (2022) noted the need for research on flexibility in algebra beyond linear equation solving. Our first contribution is to study the relationship between flexibility in the two distinct, albeit closely related, domains of linear and quadratic equation solving at the university level. We expect that most of the tasks will be solved correctly but that there will nevertheless be considerable variation in flexibility between participants.

RQ1. How flexible are university students in choosing strategies for solving linear and quadratic equations and to what extent is this choice spontaneous?

We also address the relationship between strategy flexibility and mathematics achievement highlighted by Verschaffel (2024).

RQ2. Is there a relation between university students' strategy flexibility and mathematics achievement in a course exam?

Finally, we examine differences in strategy flexibility based on students' background and grade-level.

RQ3. Does strategy flexibility differ between university students in different degree programs and between university and high-school students?

We approach these questions with a quantitative correlational research method, which is described next. Since strategy flexibility in linear and quadratic equation solving is the only type of flexibility we deal with in this article, we will mostly refer to it simply as flexibility in what follows.

Methods

Participants and context

A tri-phase test was administered at the beginning of an introductory mathematics course in the fall 2022 at a Finnish university and 306 students consented to the research-use of their data. The main dataset consists of results from this test and course exam scores. Students received additional course credit for completing the tri-phase test (regardless of giving research consent or not). The majority of the invited students agreed to participate in the study and submitted valid responses. Participants represented different degree programs, mainly computer science, mathematics and statistics, and economics (see Figure 1). In terms of prior mathematics experience, 238 students had completed the advanced mathematics track in high-school, 40 the basic track, and 25 had some other secondary-level mathematics background.

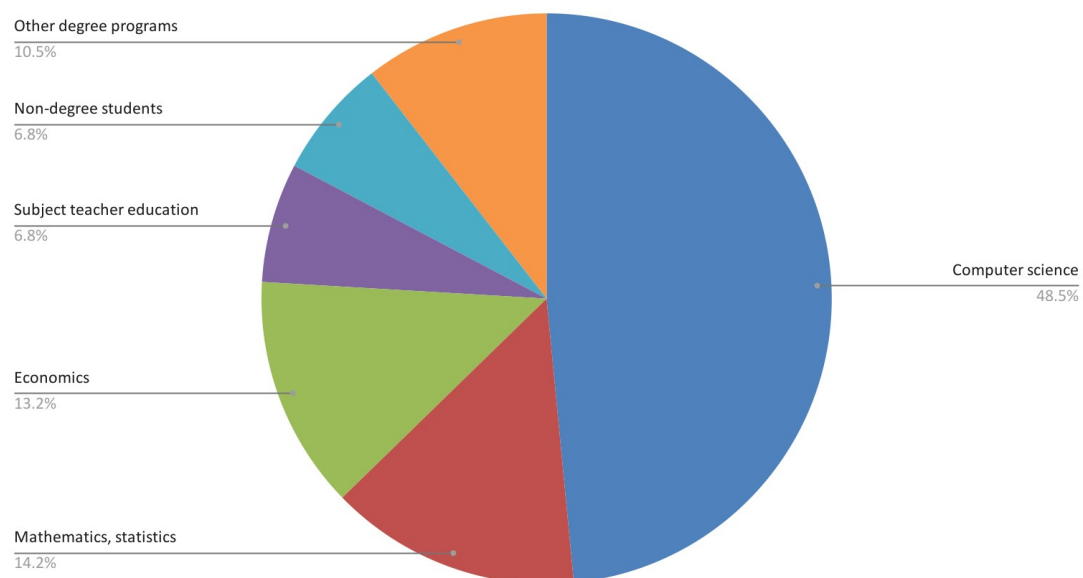


Figure 1. Participants' degree programs

The course lasted seven weeks with four hours of lectures per week and a weekly exercise set. It covered proof by induction, proof by contradiction, sets, subsets, and their size and cardinality, functions and relations and their properties, such as bijectivity, injectivity, surjectivity, and equivalence relations, the basics of logic, as well as basic operations with complex numbers, absolute value, and different representations of complex numbers. The exam described in the next subsection should give an

indication of the level at which the topics were covered. Only 260 students out of 306 participated in the course exam about two months after the tri-phase test. Some of the 46 students who did not take the exam may have completed the course by other means whereas others dropped out completely.

In addition to the main data on university students, we use data on high-school students from an open-access dataset by Jiang et al. (2022). We included only the subset of their data concerning Finnish high-school students from the basic (n=61) and advanced (n=103) mathematics tracks. The Finnish data was collected in the spring 2016 as part of a larger project (see Star et al., 2022). The secondary school system is divided into vocational schools and high schools, with about half the age cohort in each. In high school, mathematics is compulsory either at the basic level (6-8 courses) or advanced level (10-13 courses). To graduate, a student must complete 75 courses and complete a "matriculation exam" in at least four subjects not necessarily including mathematics. In the matriculation exam, basic and advanced mathematics exams are taken by roughly a third of all students each.

Materials and test procedure

We used a tri-phase test that yields measures for flexibility and spontaneous flexibility highlighting the aspects of ability and tendency from criterion (SF2), respectively (Jiang et al., 2023; Star et al., 2022). The tri-phase method also yields a measure of accuracy (ability to achieve the correct solution), which is a priori unrelated to flexibility but has been found to correlate with it (Garcia Coppersmith & Star, 2022).

As the name suggests, there are three phases in the tri-phase test. In Phase 1, students solved all six tasks without special instructions (i.e. provided one solution per task). Information about the next phase was given only after Phase 1 was completed. In particular, Phase 1 gave no hint that the objective of the test was related to flexibility and looked like a conventional equation solving test. In Phase 2, students were instructed to generate up to two additional solutions for each task. In Phase 3, students were asked to circle "the best" of their solutions for each task. These three phases are identical to those used by Xu et al. (2017), Star et al. (2022) and others, and constitute the "tri-phase method" mentioned in the introduction.

This research was preceded by two pilot studies on strategy flexibility at the university level. First, we piloted our instrument and data collection method (Ernvall-Hytönen et al., 2022). Based on the findings, we revised our test and procedures, for instance by increasing the number of flexibility tasks from 4 to 6 and by swapping a system of equations for a second quadratic equation. Tasks 1-4 were chosen from an earlier 12-item test (Jiang et al., 2023; Star et al., 2022). The original test included also items with integer and decimal numbers as well as some items with variables on both sides of the equal-sign. The items chosen were among the most difficult items from the previous studies, since the easier items might have been too easy for university students to give relevant data. The quadratic items were constructed to enable known alternative strategies (based on the zero-product rule) to the quadratic formula to be used. The equation-type $2x^2=32$ was deemed to be simple to be an effective item. The resulting tasks are shown in Table 1 on the left. We then studied the validity, especially internal consistency, of the revised test and concluded it to be good (Ernvall-Hytönen et al., in press). We refer to the latter paper for details regarding the test procedure.

Table 1. The test items with a succinct description of the generic and situational strategies for each task (Ernvall-Hytönen et al., in press).

Task	Generic	Situational
1. $4\left(x + \frac{3}{5}\right) = 12$	Distribute	Divide by 4
2. $5\left(x + \frac{3}{7}\right) + 3\left(x + \frac{3}{7}\right) = 16$	Distribute	Combine like terms
3. $\frac{2x - 6}{2} + \frac{6x - 18}{3} = 5$	G1. Expand with 3 and 2 G2. Multiply by 6	S1. Perform divisions S2. Take $2x - 6$ as common factor
4. $\frac{5x - 5}{5} + \frac{6x - 6}{6} = 5$	G1. Expand with 5 and 6 G2. Multiply by 30	S1. Perform divisions S2. Take $x - 1$ as common factor
5. $2x^2 + 4x = 0$	G1. Use quadratic formula G2. Complete the square	S1. Use zero-product property S2. Divide by x
6. $(x - 2)(x + 3) = 0$	G1. Use quadratic formula G2. Complete the square	Use zero-product property

In RQ3, we compare university students' flexibility with that of high-school students. The dataset of Jiang et al. (2022) includes student responses in a tri-phase setup to 12 tasks featuring linear equations, all of which are similar to tasks 1-4 in our test. To maximize comparability, we use only data from the four tasks that are common to the two studies, i.e. our tasks 1, 2, 3 and 4 (corresponding to their tasks 3, 5, 10 and 12, respectively). We remark that task 4 differs slightly in our test: the original equation was $\frac{5x - 5}{5} + \frac{6x - 6}{6} = 6$ while we had 5 on the right-hand side instead of 6. This does not affect the mathematical structure of the equation or the coding of strategies.

The paper-and-pencil, two-hour course exam was taken the week after the course ended. Calculators or formula collections were not allowed. The exam consisted of the following four tasks related to proof by induction, properties of relations, behavior of functions, and complex numbers.

1. Show that $(1+x)^n \geq 1+nx$ for all natural n and $x > -1$.
2. We define $x \sim y$ on the real number to mean that $x - y$ is rational. Show that \sim is an equivalence relation and determine the equivalence class of $\frac{7}{13}$.
3. (a) Give an example of a function $f: \mathbb{N} \rightarrow \mathbb{Z}$ whose converse relation is not a function from integers to natural numbers.
(b) Give an example of a function $f: \mathbb{N} \rightarrow \mathbb{Z}$ which is a surjection but not an injection.
4. (a) Show that $\overline{z+w} = \overline{z} + \overline{w}$ and $\overline{zw} = \overline{z} \overline{w}$ for complex numbers $z, w \in \mathbb{C}$.
(b) Calculate the real and imaginary parts of $\frac{3+i}{6-3i}$.

Coding and data analysis

Coding of responses was carried out independently by co-author Parikka and a research assistant. Each solution was coded in terms of accuracy as correct (1) or incorrect (0) and a strategy was identified based on the first step(s) of the solution irrespective of correctness. The strategy of each solution was categorized as situational or generic if it fits the criteria in Table 1 and as "other" if not. For instance for task 2, the generic strategy in the table is "distribute" which means that the first step typically was

$5x + \frac{15}{7} + 3x + \frac{9}{7} = 16$; the situational strategy "combine like terms" typically started with

$8\left(x + \frac{3}{7}\right) = 16$. We refer to our earlier paper for examples related to the other tasks

(Ernvall-Hytönen et al., in press). The two strategies can most of the time be distinguished even in the presence of calculation errors, which allows us to consider strategy-choice and accuracy separately. The selection of the "best" solution for a task was only considered if there were two or more solutions and ignored otherwise. The 1652 solutions of the first 151 participants were originally graded completely independently without any discussions. One or both coders marked 8.5% of strategies as unclear and 6.7% of the correctness. Of the remainder, the coders agreed on 96.4% of the strategies and 96.6% of the correctness. The unclear cases and the disagreements were discussed and resolved with the aid of co-author Ernvall-Hytönen.

We followed the general scheme of earlier papers (e.g., Star et al., 2022; Xu et al., 2017) but in some cases recognized more variety in generic or situational strategies due to the more complex situation of quadratic equations. For instance, in Task 5 the generic strategy was either using the quadratic formula (G1 in Table 1) or completing the square (G2) whereas the situational strategy was either to take x as a common factor and use the zero-product property (S1) or to divide the equation by x (S2). We refer to our earlier paper (Ernvall-Hytönen et al., in press) for more details and justifications.

Each task's response was coded for flexibility as 1, 0.5 or 0 based on the following coding scheme. A flexibility score of 1 was given if solutions to a task include a situational solution chosen as best and a generic solution, which is in line with previous research (e.g., Star et al., 2022; Xu et al., 2017). The intermediate score of 0.5 was given under two circumstances:

- (1) the solutions to a task include a situational solution chosen as best and 1-2 additional solutions that are not generic, or
- (2) the solutions to a task include both generic and situational solutions, but the situational solution is not chosen as best.

A score of 0.5 aligns closely with the concept of potential flexibility of Star et al. (2022). In all other cases, a score of 0 was given. We have shown that this scoring is closely related to earlier versions without the 0.5-category, but slightly better behaved in some cases (Ernvall-Hytönen et al., in press).

Following the coding by Star et al. (2022), the spontaneous flexibility score of a task is equal to the flexibility score if the solution in Phase 1 was situational and 0 otherwise. Accuracy of a task is scored as 1 if the Phase 1 solution attempt is correct and 0 otherwise as in prior research (Jiang et al., 2023; Liu et al., 2018; Star et al., 2022; Xu et al., 2017).

The student's flexibility, spontaneous flexibility and accuracy scores are defined as the sum of the respective scores of the six tasks. All three variables take values

between 0 and 6. Previously, we found that our six-item tri-phase test had good internal consistency with a Cronbach alpha of 0.849. Also, the flexibility subscales for linear equations (Tasks 1-4) and quadratic equations (Tasks 5&6) were internally consistent with Cronbach alphas 0.795 and 0.805, respectively (Ernvall-Hytönen et al., in press).

For comparison of groups we opted for Kruskal-Wallis and Mann-Whitney U tests. For within-group analyses, we used the Wilcoxon signed rank test. These tests were chosen because they are non-parametric, so we did not have to assume normal distribution. For basic statistical analyses, we used Spearman correlation, linear regression and Fisher's exact test. The analyses were performed using SPSS version 27.

Results

We begin this section by examining university students' flexibility when solving linear and quadratic equations. We then explore the relation between flexibility and mathematics achievement. Finally, we compare the accuracy and flexibility between groups based on participants' mathematical background and degree programs as well as between university and high-school students.

Flexibility in linear and quadratic equations

We start with some descriptive statistics. Table 2 contains averages of our variables from the tri-phase test by task, and in total. We see that the accuracy score is roughly similar in tasks 2-6 and slightly higher in the first task. Similarly, flexibility is highest in the first task, and all the other values are relatively close to one another. Spontaneous flexibility behaves somewhat differently with the highest scores in the last three tasks. The number of solutions generally decreased from task one to task six, being the lowest for quadratic equations.

Table 2. Averages by task ($n=306$).

Variable	T1	T2	T3	T4	T5	T6	Total
Accuracy	0.88	0.82	0.80	0.84	0.78	0.79	4.91
Flexibility	0.70	0.52	0.58	0.59	0.50	0.55	3.45
Spontaneous flexibility	0.35	0.29	0.31	0.43	0.43	0.42	2.23
Number of solutions	1.95	1.84	1.86	1.80	1.77	1.64	10.86

Note. Quadratic equations (task 5 & 6) highlighted

The Spearman correlation between flexibility and spontaneous flexibility is 0.789 ($p<0.001$) indicating a strong relation. We next cross-tabulated these variables as shown in Table 3. To aid in the interpretation, the scores are blocked into four groups of roughly the same size with respect to the flexibility-variable. The variables' definitions imply that spontaneous flexibility is at most equal to flexibility but never greater. However, the cross-tabulation in Table 3 shows that there are also rather few students with high flexibility and low spontaneous flexibility, and most participants are around the diagonal, as expected with the high correlation.

Table 3. Cross-tabulation of spontaneous flexibility and flexibility ($n=306$).

Spontaneous Flexibility	Flexibility				Total
	0-1.5	2-3.5	4-5.5	6	
0-1.5	69	46	15	6	136
2-3.5		38	51	9	98
4-5.5			30	24	54
6				18	18
Total	69	84	96	57	306

We next consider the linear (tasks 1-4) and quadratic subscales (tasks 5 & 6). The correlation between these subscales was 0.627 ($p<0.001$) for flexibility and 0.456 ($p<0.001$) for spontaneous flexibility. The average flexibility score was 0.60 for linear equations and 0.53 for quadratic equations, as indicated in Table 2, and the difference is statistically significant ($p<0.001$ in a related-samples Wilcoxon signed rank test). For spontaneous flexibility, the means were 0.35 and 0.42, respectively, again differing statistically ($p=0.002$ by the same test). Table 4 shows a cross-tabulation of the linear and quadratic equations subscales separately for flexibility and spontaneous flexibility.

Table 4a. Cross-tabulation of flexibility scores in quadratic and linear equations.

quadratic	linear			Total
	0-1	1.5-2.5	3-4	
0-0.5	59	36	26	121
1-1.5	11	26	46	83
2	4	18	80	102
Total	74	80	152	306

Table 4b. Cross-tabulation of spontaneous flexibility scores in quadratic and linear equations.

quadratic	linear			Total
	0-1	1.5-2.5	3-4	
0-0.5	105	41	9	155
1-1.5	36	25	13	74
2	22	23	32	77
Total	163	89	54	306

In Table 4 a) the two most common cases are to have high flexibility in both linear and quadratic equations (80 participants, i.e. 26%) and low flexibility in both (59 participants, i.e. 19%). Furthermore, more participants were above the diagonal (35%) than below (11%). In other words, it was uncommon to show flexibility in quadratic equations but not in linear ones.

Table 4 b) shows a somewhat different picture, due in part to the overall lower levels of spontaneous flexibility. To have similar levels in both linear and quadratic

equations is again most common. In particular, having low scores in both is prominent (105 participants, i.e. 34%). High spontaneous flexibility in linear equations is also strongly related to higher spontaneous flexibility in quadratic equations, with 32 out of 54 (60%) being also in the high quadratic group. However, now the proportion of participants above and below the diagonal is virtually identical (21% vs 26%).

To summarize the cross-tabulation, it was rare to be flexible in quadratic equations without being flexible in linear equations, but for spontaneous flexibility there was more variation. For instance, in the group with highest flexibility in quadratic equations only 4% had low flexibility in linear equations whereas the corresponding number for spontaneous flexibility was 29%.

Relation between flexibility and mathematics achievement

In this subsection, we study the relation between flexibility and mathematics achievement measured by a university mathematics exam. The tri-phase test yields two measures of flexibility as well as a measure of accuracy, which is more closely related to procedural fluency. We approach the relation between flexibility and achievement by measuring how well exam performance is predicted by accuracy, flexibility and spontaneous flexibility. This allows us to ensure that the exam score is specifically related to flexibility rather than equation solving skill in general.

We start with an overview of the second component of the main dataset, the course exam. The average score was 14.22, the median 15 and the mode 21, out of a maximum of 24. We compared tri-phase scores of students who participated in the exam (n=260) with those that did not (n=46). According to the Mann-Whitney U-test, the groups did not differ statistically significantly from one another: for accuracy 4.94 vs 4.76 (p=0.814), for flexibility 3.52 vs 3.09 (p=0.227) and for spontaneous flexibility 2.31 vs 1.85 (p=0.172). We conclude that participants who completed the course exam seem to be representative of all test participants.

For an initial view on the relation between the variables, we consider the correlations between exam score, accuracy, flexibility and spontaneous flexibility, which are displayed in Table 5. We see that exam score has a stronger correlation with flexibility and spontaneous flexibility than with accuracy.

Table 5. Spearman correlations between exam score, accuracy, flexibility and spontaneous flexibility.

Variable	1	2	3	4
1. Exam Score	-			
2. Accuracy	0.322	-		
3. Flexibility	0.538	0.509	-	
4. Spontaneous Flexibility	0.541	0.444	0.789	-

Note. The number of participants is 306 for every correlation not involving exam score, where it is 260. All correlations are significant at the 0.001 level.

We next investigate how the exam score can be predicted by a combination of accuracy, flexibility and spontaneous flexibility. A one-sample Kolmogorov-Smirnov test was conducted to assess the normality of the data. The results indicated that the data were not normally distributed, $D(259) = .064, p = .012$. Therefore, the test statistics of the linear regression are not reliable. Our test suggested that of the three independent

variables, flexibility and spontaneous flexibility had reasonably strong relations but accuracy did not. Therefore, accuracy was dropped, resulting in the linear regression estimate

$$\text{exam score} = 8.979 + 0.766 * \text{flexibility} + 1.104 * (\text{spontaneous flexibility}),$$

which accounted for 30.6% of the total variance. Note that this part of the linear regression is not contingent on the normality assumption.

This regression only captures linear relationships and it does not provide reliable test statistics due to the non-normality of our data. To investigate possible non-linear relationships between the variables and their statistical significance, we created four different groups based on exam score: the ranges 0-9, 10-14, 15-19 and 20-24 divide students into four roughly equal size quartiles (with 65, 58, 70 and 67 students, respectively). The average accuracy, flexibility and spontaneous flexibility scores of these groups are shown in Figure 2.

The differences between groups are statistically significant in each of the three variables (three separate Kruskal-Wallis tests, $p < 0.001$). We performed post-hoc Mann-Whitney U tests to compare the groups pairwise; the results are summarized in Figure 2. In flexibility and spontaneous flexibility, all groups except the lowest two were different ($p < 0.001$). In accuracy, only the highest performing group was statistically different from the other groups ($p < 0.001$) while no other differences were statistically significant. In other words, the top quartile in the exam outperformed the other quartiles in accuracy, flexibility as well as spontaneous flexibility. This was the only significant effect concerning accuracy and the top quartile had very high accuracy scores. Furthermore, the second quartile in the exam outperformed the lower quartiles in flexibility and spontaneous flexibility but not accuracy. There were no significant differences in any variable between the 3rd and 4th quartiles.

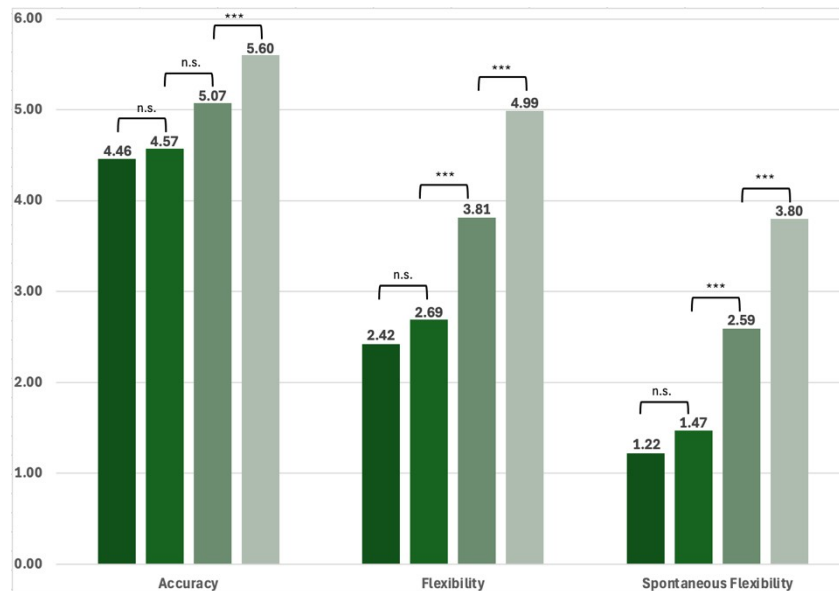


Figure 2. Comparison of group averages. The columns, from the left, correspond to exam scores 0-9, 10-14, 15-19 and 20-24 with 65, 58, 70 and 67 participants, respectively. Differences of consecutive groups are shown as non-significant (n.s.) or significant at the $p < 0.001$ level (***)

For even more nuanced insight, we cross-tabulated the exam score groups with accuracy, flexibility and spontaneous flexibility ranges. Note that the accuracy groups are skewed in size, whereas the other three variables roughly correspond to quartiles.

Table 6. Cross-tabulation of tri-phase test variables and exam score groups.

Exam score	n	Accuracy			Flexibility				Spontaneous Flexibility			
		0-3	4-5	6	0-1.5	2-3.5	4-5.5	6	0-0.5	1-1.5	2-3.5	4-6
0-9	65	14	27	24	28	22	11	4	31	12	19	3
10-14	58	10	27	21	17	22	17	2	15	23	18	2
15-19	70	10	27	33	8	20	32	10	11	11	28	20
20-24	67	1	17	49	3	10	20	34	4	7	16	40
Total	260	35	98	127	54	74	82	50	61	53	81	65

From Table 6, we can make the following observations: the weak relation between accuracy and exam score is a consequence of a large proportion of the students obtaining the maximum score in accuracy while at the same time exhibiting the full range of exam scores. For both flexibility and spontaneous flexibility, the values are centered around the diagonal (with 44.6% and 46.9% on the diagonal) indicating a high correlation. The major difference between flexibility and spontaneous flexibility lies in the different ranges that constitute the quartiles; for instance, the top quartile in spontaneous flexibility contains scores 4-6 whereas flexibility only includes the score of 6. This means that spontaneous flexibility is more fine-tuned in the high flexibility case and flexibility is more sensitive in the low flexibility case. For instance, out of the 17 participants with spontaneous flexibility score of 6, all but one (i.e. 94.1%) were in the top half of exam scores whereas the corresponding percentage for flexibility was 64.0%. The complementary nature of flexibility and spontaneous flexibility is reflected in both being included in the linear regression model, above.

Relation between flexibility and student background

In this subsection, we first consider university students' high-school mathematics background. Averages of accuracy, flexibility, spontaneous flexibility and exam score are shown in Table 7. We find that each variable is higher in the group with advanced mathematics background (four Kruskal-Wallis tests, $p < 0.001$). We note that the drop-out rates appear to be different, namely 13.5% and 27.5% for advanced and basic mathematics backgrounds ($p = 0.0323$ according to Fisher's exact test).

Table 7. Average accuracy, flexibility, spontaneous flexibility and exam score by high school mathematics background.

Background	n	Accuracy	Flexibility	Spont. Flex.	Exam score
High school advanced mathematics	238/206	5.17	3.84	2.49	14.90
High school basic mathematics	40/29	3.68	1.94	1.09	10.24

Note. The second number in the n-column pertains to exam score while the first number concerns the other three variables.

We then consider different degree programs. As shown in Figure 1, three of the degree program categories were quite small, so Table 8 shows averages only for the three largest groups. Drop-out rates are not significantly different between degree programs (pairwise Fisher exact tests, $p=0.24$ or higher).

Table 8. Average accuracy, flexibility, spontaneous flexibility and exam score by degree program.

Degree program	n	Accuracy	Flexibility	Spont. Flex.	Exam score
Computer Science	143/118	4.60	2.93	1.87	12.87
Economics	41/37	5.10	3.48	1.95	11.84
Mathematics & statistics	45/41	5.33	4.58	3.23	17.83

Note. The second number in the n-column pertains to exam score while the first number concerns the other three variables.

Kruskal-Wallis tests show that the differences between groups are statistically significant in all four variables ($p=0.019$ for accuracy and $p<0.001$ for the other three). Post-hoc comparisons with the Mann-Whitney U test showed significant differences between mathematics and computer science in all four variables ($p=0.010$ for Accuracy and $p<0.001$ for the others). Differences between mathematics and economics students were significant in all variables except accuracy ($p<0.001$ for exam score, $p=0.010$ for flexibility and $p=0.002$ for spontaneous flexibility). Between computer science and economics students, the differences were not significant in any variable. We also investigated differences between degree programs in flexibility separately for linear and quadratic equations, but we did not find any effect of equation type.

Since different degree programs may have different background profiles, we next investigate whether this explains the differences between degree programs. For example, in the mathematics degree program, all but one student had an advanced mathematics high-school background. To control for the impact of high-school background, we calculated the average scores for each degree program considering only students with advanced mathematics background from high school. The other background alternatives were so uncommon that they were not considered. The results are shown in Table 9.

Table 9. Average accuracy, flexibility, spontaneous flexibility and exam score by degree program for students with advanced mathematics background from high school.

Degree program	n	Accurac y	Flexibilit y	Spont. Flex.	Exam score
Computer Science	100/85	5.00	3.50	2.21	13.94
Economics	36/33	5.25	3.64	2.00	12.15
Mathematics & statistics	44/40	5.32	4.58	3.20	17.90

Note. The second number in the n-column pertains to exam score while the first number concerns the other three variables.

Kruskal-Wallis tests show statistically significant differences between groups in flexibility ($p=0.003$), spontaneous flexibility ($p=0.006$) and exam score ($p<0.001$) but not accuracy ($p=0.305$). Post-hoc comparisons with the Mann-Whitney U test showed no differences between computer science and economics whereas both groups differed significantly from mathematics in flexibility, spontaneous flexibility and exam score ($p<0.005$, except for flexibility between mathematics and economics students with $p=0.026$).

Comparing Tables 8 and 9 and the statistical analyses, we see that the different prevalence of high-school tracks in the degree programs explains the difference in accuracy between mathematics and computer sciences. Considering the high-school background does not change the statistical significance of differences between groups for any other variable.

We next compare the flexibility of university students from our test with that of high school students from the open access dataset by Jiang et al. (2022) described earlier. Table 10 shows the averages in the four tasks common to the datasets.

Table 10. Average total score from tasks with linear equations (tasks 1-4).

Level	Group	n	Accur acy	Flexibilit y	Spont. Flex.
University students	All	306	3.34	2.40	1.39
	Advanced mathematics background	238	3.47	2.65	1.52
	Basic mathematics background	40	2.63	1.44	0.76
High-school students	All	164	1.58	1.08	0.33
	Advanced mathematics track	103	2.18	1.55	0.50
	Basic mathematics track	61	0.56	0.30	0.03

We first compared all university students with all high school students (e.g., accuracy 3.34 vs 1.58). University students with a high school advanced mathematics background were then compared with high school students studying advanced mathematics (e.g., flexibility 2.65 vs 1.55), and the same for basic mathematics. All the differences were significant at the 0.001-level according to Mann-Whitney U tests. Specifically, university students with an advanced mathematics background

outperformed high-school students from the advanced mathematics track in accuracy, flexibility and spontaneous flexibility; the same was true for the basic mathematics background compared to the basic mathematics track.

Discussion

Flexibility in linear and quadratic equations

The first objective of this study was to examine the flexibility of university students, in particular regarding spontaneity as well as differences between linear and quadratic equations (RQ1). We found overall quite high levels of flexibility with 50% of participants getting at least 4 out of 6 as the flexibility score. However, only 23.5% of participants got at least 4 as the spontaneous flexibility score. The share of spontaneous flexibility in relation to non-spontaneous flexibility is similar to earlier research (Star et al., 2022) although the absolute percentages are higher.

We examined flexibility using both linear and quadratic equations. The flexibility profiles between linear and quadratic equations differed from one another, but also other features of equations like complexity played a role. Our data shows that it is rare to show high flexibility in quadratic equations with low flexibility in linear equations (only 4% of participants). While still not the norm, this was much more common for spontaneous flexibility (29%). We offer two possible explanations for this. It seems that if the technical demands, e.g. the number of steps and calculations, of the generic strategy are lower, then there is less incentive to use the situational strategy. Conversely, higher technical demands, as in task 4, lead to more situational strategies. This may explain the very high flexibility score and the relatively low spontaneous flexibility score in task 1, as well as the high spontaneous flexibility in task 4. Another explanation is related to quadratic equations, where the situational strategy to solve quadratic equations is explicitly taught alongside the generic one in Finnish schools whereas more emphasis is on the generic strategy (i.e. the standard algorithm) for solving linear equations (FNBE, 2014). That could make recognizing the applicability of the generic strategy the largest hurdle to flexibility for some students (cf. the discussion on practical flexibility by Xu et al., 2017 and Liu et al., 2018).

Relation between flexibility and mathematics achievement

We then turn to RQ2. Our data showed a reliable and moderately strong relationship between strategy flexibility in solving elementary algebraic equations and achievement in an exam covering diverse topics like induction, relations, functions and complex numbers without any direct link to equation solving. It seems that strategy flexibility can be a good predictor of mathematics achievement across mathematics topics. This is particularly interesting since university mathematics courses require skills much beyond simple equation solving abilities. It is possible that even better results might be obtained by extending the difficulty range of the flexibility-tasks, for instance by considering arithmetic and differentiation tasks alongside algebraic tasks, but this is left for future research. Furthermore, we found some indications that spontaneous flexibility may function as a more demanding version of (non-spontaneous) flexibility at least when predicting mathematics achievement; it would be interesting to explore this topic further. Another topic to study is individual development to see if flexibility becomes spontaneous with sufficient practice or if it is more of an individual preference where some people go the spontaneous flexibility route and others the (non-spontaneous) flexibility route.

In a linear regression model with the three tri-phase test variables (accuracy, flexibility and spontaneous flexibility) as predictors, accuracy turned out to be

redundant and a two variable prediction model based on only flexibility and spontaneous flexibility accounted for 30% of the variation in exam score. This result is broadly consistent with earlier results (e.g., Star et al., 2022; Torbeyns et al., 2017) showing a positive relationship between strategy flexibility and mathematics achievement. However, our result differs from the finding of Hästö et al. (2019) that accuracy, rather than flexibility, was the better predictor for mathematics achievement in a national exam. A possible explanation for this discrepancy is that Hästö et al. (2019) observed that higher flexibility was related to a higher probability of participating in the advanced mathematics exam: taking a more demanding exam is also a form of mathematical achievement, which is, however, not comparable to exam scores. Another possible explanation is based on the levels of variables: in the high-school sample, flexibility was quite low and the flexibility variable may have had a floor-effect whereas the accuracy variable in the university sample was correspondingly high and subject to a ceiling-effect. Such cut-off effects would make flexibility a better predictor in groups with high levels of flexibility and accuracy in groups with low levels of flexibility.

Regarding non-linear relations between exam score and tri-phase test variables, we found that there were no significant differences in any variable between the 3rd and 4th exam-score quartile. This supports the idea that there is a threshold that needs to be reached before flexibility becomes a good predictor of mathematics achievement. The result concerning accuracy qualifies the ceiling-effect mentioned above: there is significant clumping with accuracy averages in the range 4.57-5.07 for the 2nd-4th quartiles but only the top quartile is subject to a proper ceiling-effect at an average of 5.60.

Cut-off effects may also contribute to interpreting findings of Bye et al. (2022) and Shaw et al. (2020) that flexibility and mathematics achievement were not related. We suggested that the participants need to score relatively high on flexibility for it to be a good predictor of mathematics achievement. The tasks of Bye et al. (2022) may have been too routine for flexibility to predict achievement: their measure of flexibility is based on reaction times related to arithmetic facts like $9-7=2$ and $2+7=9$ and corresponding simple equations like $x-7=2$; this measure was found not to predict mathematics achievement of university students and our proposal is that this may be because the tasks were too easy mathematically. A lack of sufficient mathematical challenge may also be relevant to interpreting the absence of predictive power of flexibility found by Shaw et al. (2020), although it is not as closely related as they predicted affective and cognitive variables rather than achievement.

Relation between flexibility and student background

Finally, we compared the flexibility of different groups of students in connection to RQ3. Since the data was collected early in a first period course that many students take in their first year, we did not expect a large effect of university coursework, but the different degree programs might attract students with different profiles. Perhaps unsurprisingly, students in the mathematics and statistics degree programs outperformed their peers in flexibility, spontaneous flexibility and exam score (Table 8) but there were no significant differences between computer science and economics. Participants' high school background turned out to be more significant than the degree program.

Indeed, participants with advanced mathematics high-school background showed over twice the spontaneous flexibility score and almost twice the flexibility score in our test compared to those with basic mathematics background. It is worth noting that since most students have had little university mathematics studies prior to

completing the tri-phase test, they mainly solved the tasks based on their earlier mathematics knowledge from high-school. This suggests that the advanced track in high school gives a significantly better foundation to develop flexibility, at least in equation solving. Additionally, being flexible is known to be related to choosing and persevering in the advanced mathematics track (Hästö et al., 2019). It would be interesting to study if the background plays an equally large role if flexibility were tested in a domain of university mathematics such as linear algebra.

Our results provide some support for the suggestion of Kontorovich (2020) and Broley and Hardy (2022) that flexibility could be a predictor for students' ability and tendency to use more advanced theoretical tools in linear algebra and analysis over brute-force calculation. It is worth noting that the course exam measuring mathematics achievement in our study was less advanced than what these authors described and consequently we did not analyze the exam answers from this point of view.

To answer the second part of RQ3, we also compared university and high-school students' flexibility and found indications that university students are more flexible, both spontaneously and non-spontaneously, than high-school students. Previous research (e.g. Hästö et al., 2019; Star et al., 2022) had shown that there is a substantial flexibility difference between high-school students in the advanced and basic mathematics tracks in Finland. Therefore, it is natural to consider the track when performing the comparison. With the caveats below, we found that university students with an advanced mathematics background outperformed high-school students from the advanced mathematics track in accuracy, flexibility and spontaneous flexibility; the same was true for the basic mathematics background compared to the basic mathematics track. The differences are due to at least two factors: the university students had approximately one year more of mathematics classes (in high-school) and most of them would have performed above average in high-school since they ended up in a university mathematics class (in Finnish universities, most degree programs have no mathematics requirement).

However, interpretation is challenging as the high-school sample is not representative and it was collected almost a decade before the university data. It is also worth noting that high-school students completed the four tasks as part of a 12-task test on linear equations in a timed setting in class whereas the university students completed the same tasks alongside two other tasks on quadratic equations without time limitation on their own time. Time pressure may account for part of the difference between university and high-school students; this is supported by the observation that high-school students performed better flexibility-wise in the first task of the test than in the last ones (cf. Figure 2 in Hästö & Palkki, 2019 and Figure 1 in Star et al., 2022), whereas the corresponding differences are smaller for university students (Table 2). Despite these reservations, our results give some support at the university-level to the hypothesis that higher grade-level is related to higher strategy flexibility (e.g., Torbeyns et al., 2009). While this seems like a natural conclusion, previous research has not always supported the claim (e.g., Van Der Auwera et al., 2022). One possible impediment is a ceiling after which developing flexibility becomes harder so that there would be a weaker relation with grade-level at university. However, our data suggests to the contrary that measuring flexibility with equation solving is perhaps even better suited for university students who have a better grasp of the basics.

Limitations and future research

While the group of students answering the test was sizable, a number of students either did not participate in the test or did not give permission to use their data in research.

Therefore, the study sample may not be representative of all students participating in the class. This limits the generalizability of the results within the science student population of a Finnish university; of course, generalization to the whole cohort or other countries or cultures is even more restricted.

We only measured flexibility at the beginning of the course and mathematics achievement at the end of the course. Thus, we cannot say anything about the development during the course. A longitudinal study over a longer period could be interesting to see whether flexibility develops during normal university courses. Furthermore, this could be combined with an intervention to increase flexibility. Such intervention could consist of emphasizing multiple solution strategies (Maciejewski & Star, 2016) or teaching a topic in an open-ended fashion (Star & Rittle-Johnson, 2008).

It is perhaps surprising that the differences in flexibility among students in elementary equations were still quite large. It is unclear if the differences would be smaller for instance among third year mathematics majors. To get a more comprehensive understanding of university students' flexibility, one could study flexibility in a broader context, for instance by considering calculus (e.g. Maciejewski, 2022; Maciejewski & Star, 2016) and arithmetic (e.g. Hickendorff, 2018) alongside algebra.

Conclusion

We studied the flexibility of university students using a six-item tri-phase test in the domain of linear and quadratic equations. Strategy flexibility is defined as the knowledge of multiple strategies and ability and tendency to select the most appropriate strategy for the problem at hand and our results showed variation between students in all three aspects (knowledge, ability and tendency). Most students displayed at least some flexibility although fewer did so spontaneously. Flexibility was positively related to mathematics achievement, especially when flexibility was spontaneous. Finally, both students' degree programs and their mathematical background had a strong relation to their flexibility.

Acknowledgment

We would like to thank the anonymous referees' for their useful comments.

References

- Broley, L., & Hardy, N. (2022). The evolution of students' learning from calculus to analysis: How students solve analysis tasks that look like calculus tasks. *International Journal of Research in Undergraduate Mathematics Education*, 8, 269–292. 10.1007/s40753-022-00171-2
- Bye, J.K., Harsch, R.M., & Varma, S. (2022). Decoding fact fluency and strategy flexibility in solving one-step algebra problems: An individual differences analysis. *Journal of Numerical Cognition*, 8(3), 281–294. 10.5964/jnc.7093
- Cox, S.K., Burns, M.K., Hughes, E.M., Wade, T., & Brown, M. (2024). Defining, measuring, and teaching mathematical flexibility: a content analysis of extant research. *The Elementary School Journal*, 124(3), 479–498. 10.1086/728591
- Ernvall-Hytönen, A.-M., Hästö, P., Krzywacki, H., & Parikka, S. (2022). Procedural flexibility in early university mathematics. *FMSERA Journal*, 5, 46-60.
- Ernvall-Hytönen, A.-M., Hästö, P., Krzywacki, H., & Parikka, S. (in press). Measures of strategy flexibility in an abbreviated tri-phase test on linear and quadratic equation solving. *FMSERA Journal*.

- FNBE (Finnish National Board of Education) (2014). *National Core Curriculum for Basic Education 2014*. Helsinki: Finnish National Agency for Education.
- FNBE (Finnish National Board of Education) (2019). *National Core Curriculum for General Upper Secondary Education 2019*. Helsinki: Finnish National Agency for Education.
- Garcia Coppersmith, J., & Star, J.R. (2022). A complicated relationship: examining the relationship between flexible strategy use and accuracy. *Journal of Numerical Cognition*, 8(3), 382–397. 10.5964/jnc.7601
- Heinze, A., Star, J.R. & Verschaffel, L. (2009). Flexible and adaptive use of strategies and representations in mathematics education. *ZDM Mathematics Education*, 41, 535–540. 10.1007/s11858-009-0214-4
- Hästö, P., & Palkki, R. (2019). Finnish students' flexibility and its relation to speed and accuracy in equation solving. *Nordic Studies Math. Education NOMAD*, 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. 10.30935/scimath/9530
- Hästö, P., & Star, J.R. (2024). Strategic flexibility in university mathematics. *Notices of the American Mathematical Society*, 71(7), 917–923.
- Hickendorff, M. (2018). Dutch sixth graders' use of shortcut strategies in solving multidigit arithmetic problems. *European Journal of Psychology of Education*, 33, 577–594. 10.1007/s10212-017-0357-6
- Hickendorff, M. (2022). Flexibility and adaptivity in arithmetic strategy use: What children know and what they show. *Journal of Numerical Cognition*, 8(3), 367–381. 10.5964/jnc.7277
- Hickendorff, M., McMullen, J., & Verschaffel, L. (2022). Mathematical flexibility: theoretical, methodological, and educational considerations. *Journal of Numerical Cognition*, 8(3), 326–334. 10.5964/jnc.10085
- Hong, W., Star, J.R., Liu, R.-D., Jiang, R., & Fu, X. (2023). A systematic review of mathematical flexibility: concepts, measurements, and related research. *Educational Psychology Review*, 35, Article 104. 10.1007/s10648-023-09825-2
- Huntley, M.A., Marcus, R., Kahan, J., & Miller, J.L. (2007). Investigating highschool students' reasoning strategies when they solve linear equations. *The Journal of Mathematical Behavior*, 26(2), 115–139. 10.1016/j.jmathb.2007.05.005
- Jiang, R., Star, J.R., & Hästö, P. (2022, May 16). *Procedural flexibility in mathematics in three countries*. 10.17605/OSF.IO/KWQDB
- Jiang, R., Star, J.R., Hästö, P., Li, L., Liu, R.-D., Tuomela, D., Joglar Prieto, N., Palkki, R., Abánades, M., & Pejlare, J. (2023). Which one is the “best”: a cross-national comparative study of students' strategy evaluation in equation solving. *International Journal of Science and Mathematics Education*, 21, 1127–1151. 10.1007/s10763-022-10282-6
- Keleş, T., & Yazgan, Y. (2021). Gifted eighth, ninth, tenth and eleventh graders' strategic flexibility in non-routine problem solving. *The Journal of Educational Research*, 114, 332–345. 10.1080/00220671.2021.1937913

- Kontorovich, I. (2020). Theorems or procedures? Exploring undergraduates' methods to solve routine problems in linear algebra. *Mathematics Education Research Journal*, 32, 589–605. 10.1007/s13394-019-00272-3
- Liu, R.D., Wang, J., Star, J.R., Zhen, R., Jiang, R.H., & Fu, X.C. (2018). Turning potential flexibility into flexible performance: moderating effect of self-efficacy and use of flexible cognition. *Frontiers in Psychology*, 9, Article 646. 10.3389/fpsyg.2018.00646
- Maciejewski, W. (2022). Between confidence and procedural flexibility in calculus. *International Journal of Mathematical Education in Science and Technology*, 53(7), 1733–1750. 10.1080/0020739X.2020.1840639
- Maciejewski, W., & Star, J.R. (2016). Developing flexible procedural knowledge in undergraduate calculus. *Research in Mathematics Education*, 18, 1–18. 10.1080/14794802.2016.1148626
- NCTM (National Council of Teachers of Mathematics) (2023). *Procedural fluency in mathematics*. NCTM Position Statement, January 2023.
- 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. 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. 10.1037/a0014224
- Rittle-Johnson, B., Star, J.R., & Durkin, K. (2009). The importance of prior knowledge when comparing examples: Influences on conceptual and procedural knowledge of equation solving. *Journal of Educational Psychology*, 101(4), 836–852. 10.1037/a0016026
- Rittle-Johnson B., Star J.R., & Durkin K. (2012). Developing procedural flexibility: are novices prepared to learn from comparing procedures? *British Journal of Educational Psychology*, 82(3), 436–455. 10.1111/j.2044-8279.2011.02037.x
- Shaw, S.T., Pogossian, A.A., & Ramirez, G. (2020). The mathematical flexibility of college students: The role of cognitive and affective factors. *British Journal of Educational Psychology*, 90(4), 981–996. 10.1111/bjep.12340
- Star, J.R., & Newton, K.J. (2009). The nature and development of experts' strategy flexibility for solving equations. *Zentralblatt Für Didaktik Der Mathematik*, 41, 557–567. 10.1007/s11858-009-0185-5
- Star, J.R., & Rittle-Johnson, B. (2008). Flexibility in problem solving: The case of equation solving. *Learning and Instruction*, 18, 565–579. 0.1016/j.learninstruc.2007.09.018
- Star, J.R., Tuomela, D., Joglar-Prieto, N., Hästö, P., Palkki, R., Abánades, M.Á., Peljare, J., Jiang, R. H., Lijia, L., & Liu, R. D. (2022). Exploring students' procedural flexibility in three countries. *International Journal of STEM Education*, 9, 1–18. 10.1186/s40594-021-00322-y
- Torbeyns, J., De Smedt, B., Ghesquière, P., & Verschaffel, L. (2009). Acquisition and use of shortcut strategies by traditionally schooled children. *Educational Studies in Mathematics*, 71, 1–17. 10.1007/s10649-008-9155-z

- Torbeyns, J., Hickendorff, M., & Verschaffel, L. (2017). The use of number-based versus digit-based strategies on multi-digit subtractions: 9–12-year-olds' strategy use profiles and task performances. *Learning and Individual Differences, 58*, 64–74. 10.1016/j.lindif.2017.07.004
- Van Der Auwera, S., Torbeyns, J., De Smedt, B., Verguts, G., & Verschaffel, L. (2022). The remarkably frequent, efficient, and adaptive use of the subtraction by addition strategy: A choice/nochoice study in fourth- to sixth-graders with varying mathematical achievement levels. *Learning and Individual Differences, 93*, Article 102107. 10.1016/j.lindif.2021.102107
- Verschaffel, L. (2024). Strategy flexibility in mathematics. *ZDM Mathematics Education, 56*, 115–126. 10.1007/s11858-023-01491-6
- Xu, L., Liu, R.-D., Star, J.R., Wang, J., Liu, Y., & Zhen, R. (2017). Measures of potential flexibility and practical flexibility in equation solving. *Frontiers in Psychology, 8*, article 1368. <https://doi.org/10.3389/fpsyg.2017.01368>