

A Complicated Relationship: Examining the Relationship Between Flexible Strategy Use and Accuracy

Jeannette Garcia Coppersmith¹, Jon R. Star¹

[1] *Harvard Graduate School of Education, Cambridge, MA, USA.*

Journal of Numerical Cognition, 2022, Vol. 8(3), 382–397, <https://doi.org/10.5964/jnc.7601>

Received: 2021-10-01 • Accepted: 2022-06-07 • Published (VoR): 2022-11-16

Handling Editor: Lieven Verschaffel, University of Leuven, Leuven, Belgium

Corresponding Author: Jon R. Star, 442 Gutman Library, 6 Appian Way, Cambridge, MA 02138, USA. E-mail: jon_star@harvard.edu

Related: This article is part of the JNC Special Issue “Mathematical Flexibility”, Guest Editors: Marian Hickendorff, Jake McMullen, & Lieven Verschaffel, Journal of Numerical Cognition, 8(3), <https://doi.org/10.5964/jnc.v8i3>

Abstract

This study explores student flexibility in mathematics by examining the relationship between accuracy and strategy use for solving arithmetic and algebra problems. Core to procedural flexibility is the ability to select and accurately execute the most appropriate strategy for a given problem. Yet the relationship between strategy selection and accurate execution is nuanced and poorly understood. In this paper, this relationship was examined in the context of an assessment where students were asked to complete the same problem twice using different approaches. In particular, we explored (a) the extent to which students were more accurate when selecting standard or better-than-standard strategies, (b) whether this accuracy-strategy use relationship differed depending on whether the student solved a problem for the first time or the second time, and (c) the extent to which students were more accurate when solving algebraic versus arithmetic problems. Our results indicate significant associations between accuracy and all of these aspects— we found differences in accuracy based on strategy, problem type, and a significant interaction effect between strategy and assessment part. These findings have important implications both for researchers investigating procedural flexibility as well as secondary mathematics educators who seek to promote this capacity among their students.

Keywords

flexibility, accuracy, strategy, standard algorithm, algebra

Promoting the development of students’ problem-solving abilities, particularly procedural flexibility, is one important aim in mathematics educational practice and policy (Institute of Educational Sciences, 2012; Kilpatrick et al., 2001; National Council of Teachers of Mathematics, 2000, 2014). A student’s ability to approach a mathematics problem in a variety of ways and their knowledge of when to apply a solution method to a particular problem is central to flexibility and flexible knowledge (Star, 2005, 2007; Star & Rittle-Johnson, 2008; Star & Seifert, 2006). We define procedural flexibility as a student’s knowledge of multiple approaches or strategies for solving a mathematics problem and their ability to select the most appropriate one (Star, 2005). Among the multiple strategies that can be used to solve any mathematics problem, some strategies are arguably more appropriate under particular problem-solving circumstances than others. Core to flexibility is selecting an appropriate mathematical strategy depending on specific features of the problem and the problem-solving circumstances (Star, 2005). Yet regardless of how elegant or appropriate a strategy is, the ability to implement it and solve a problem accurately is of equal importance. The purpose of this study is to examine the relationship between strategy use and accuracy. This relationship was explored in the context of an



assessment where students were asked to complete the same problem twice using different approaches, allowing us to examine the relationship between strategy use and accuracy beyond a student's primary approach of choice. Further, we examine whether the relationship between strategy use and accuracy depends on the problem type—arithmetic or algebraic—to add further nuance to our current understandings about the role of structural features and mathematical domains for the relationship between strategy selection and accuracy.

Relationships Between Strategy Appropriateness and Strategy Accuracy

Among the many dimensions along which problem-solving strategies can differ are two that are of particular interest in this study: strategy appropriateness and strategy accuracy. Strategy appropriateness lies at the core of procedural flexibility (Star, 2005). Some strategies may be better than others for a given problem by virtue of a number of factors, including efficiency of solving and elegance of the steps. Other characterizations of appropriateness consider situational variables and the learner (e.g., Verschaffel et al., 2009); here, we are primarily concerned with the strategies students employ with respect to the specific features of the task and their efficiency and elegance. While mathematicians disagree about how to objectively define elegance in their discipline, they widely agree that strategy appropriateness and elegance go hand-in-hand (Hardy, 1940). For example, within the algebraic domain of linear equation solving, many equations can be solved using a so-called standard algorithm (Buchbinder et al., 2015; Star & Seifert, 2006). Such an algorithm applies to a broad range of problems and is reasonably efficient. A standard algorithm for solving a linear equation such as $3(x + 1) = 15$ involves first distributing the coefficient 3 to obtain the expression $3x + 3$, then subtracting 3 from both sides, and finally dividing by 3 to arrive at $x = 4$ (Buchbinder et al., 2015; Star & Seifert, 2006). Similarly, a standard approach for adding a series of integers such as $146 + 12 - 46 + 88$ would be to add from left to right: $146 + 12$ is 158, $158 - 46$ is 112, and $112 + 88$ is 200.

Among other possible strategies, some are arguably better than the standard algorithm, where better (or “situationally appropriate”; Star et al., 2022) may mean that the strategy is more elegant and/or better matched to the structural features of the problem. To illustrate, for the above equation $3(x + 1) = 15$, an arguably better strategy would involve dividing both sides of the equation by 3 as a first step. This approach is considered better than the standard algorithm due to the affordances created by having 3 be a factor of 15—a structural feature of the problem—and to the fewer steps required to solve it. In the case of the addition problem $146 + 12 - 46 + 88$, an arguably better approach would be to recognize a structural, numerical relationship between the to-be-added numbers (that both $146 - 46$ and $12 + 88$ are easy to compute, as is their sum) and to add the non-consecutive pairs of integers to take advantage of this. Furthermore, recognizing and taking advantage of structural features of a problem when selecting a strategy also illustrates linkages between flexibility and conceptual knowledge (e.g., Schneider et al., 2011). In the former example, use of the better strategy may imply a more sophisticated understanding of the concept of variable, in treating $(x + 1)$ as a variable term. And in the later example, use of the more situational appropriate strategy relies both upon the conceptual principle of commutativity as well as the relationship between subtraction and addition.

But is there a general relationship between strategy appropriateness and strategy accuracy? In other words, which strategies tend to be more accurately implemented by students, those that are standard algorithms or those that are more situationally appropriate? On the one hand, an argument can be made that a standard approach tends to be the strategy most related to accuracy. Standard approaches are by definition broadly applicable to a wide range of problems. As a result, such algorithms can be automatically executed without a great deal of attention to the specific structural features of a problem. Such routine execution of standard algorithms can be efficient and reduce the likelihood of error. Prior research has suggested a “freed resources” mechanism behind this potential benefit of standard algorithms, where highly routinized strategies enable students to focus on the relationships and operations in a problem with greater facility (Kotovsky et al., 1985; McNeil & Alibali, 2004; Shiffrin & Schneider, 1977; Shrager & Siegler, 1998). In addition, well-rehearsed standard approaches may also lead to greater confidence and trust among users of these algorithms, which may also be linked to higher accuracy. Finally, use of a standard algorithm may place a lower burden on working memory capacity during anxiety-inducing assessments, enabling students to perform more successfully as compared to students who use other strategies (Torbeys & Verschaffel, 2013, 2016).

But on the other hand, one might hypothesize that strategies that are more appropriate or better than the standard algorithm would result in greater accuracy. Situationally appropriate strategies take advantage of structural features of problems and tend to have fewer operations, and this may reduce the opportunity and likelihood for errors. When students opt for a strategy that departs from a highly routinized, standard approach, they may engage more carefully in on-the-spot encoding of problem features (McNeil & Alibali, 2004); this break from automaticity, as well as the more conscious and deliberate attention to the problem-solving process, may increase accuracy. Further, highly practiced strategies such as standard algorithms cause students to only encode the features of the problem necessary for executing their strategy (McNeil & Alibali, 2004), and this may cause students to miss opportunities for structural efficiencies and shortcuts available to them. For these reasons, better-than-standard approaches may lead to greater success in problem solving.

As an alternative to either of these hypotheses, it may instead be the case that a general relationship does not exist between strategy appropriateness and strategy accuracy – but rather, that this relationship is an interaction related to the structural features of a given problem. In particular, for problems where it is relatively straightforward to identify an alternative, better strategy (such as in the specific linear equation and integer addition examples previously shown), perhaps the use of the more situationally appropriate strategy leads to greater accuracy, for the reasons noted above. As another example, consider the fraction addition problem $18/36 + 21/42$ (Newton, 2008), where there may be considerable advantage in terms of accuracy in recognizing that both fractions are equivalent to $1/2$ rather than using the standard algorithm (such problems have been termed “flexibility-eligible” problems; Hästö et al., 2019). But for problems where it may not be readily apparent whether there is a better alternative to the standard algorithm, perhaps the standard algorithm holds greater promise for accuracy. For example, consider the linear equation $3(x - 4) - 15x = -72$. Alternative strategies may exist – some of which might be considered better than the standard approach, such as dividing each term by 3 as a first step – but given the greater effort required to generate and execute these alternatives, the standard approach may be more likely to lead to the correct solution.

Other Considerations for the Relationship Between Accuracy and Strategy Use

An additional consideration with regard to the relationship between strategy appropriateness and strategy accuracy concerns the particulars of how these constructs are assessed. Specifically, in prior studies of procedural flexibility (e.g., Star & Seifert, 2006), a common means for assessing strategy appropriateness is to ask students to re-solve a previously completed problem but using a different approach. Inferences about students’ flexibility are then made based on whether a better-than-standard or a standard strategy was used in the student’s first solution attempt or in subsequent attempts (e.g., Xu et al., 2017), where typically the use of a better-than-standard approach in a student’s first attempt is viewed as an indicator of more sophisticated procedural flexibility. Using this same logic, it is not clear whether students would exhibit greater accuracy in their first attempt or in their second attempt on a problem. Students’ first attempt will presumably draw upon the strategy that they are most comfortable and familiar with, which could produce greater accuracy. But alternatively, after having solved the problem once, the students’ greater familiarity with the problem (and its solution) could result in greater accuracy on the second attempt. Furthermore, there could conceivably be an interaction between the section of the assessment and the strategy used, whereby (for example) the first attempt strategy is more accurate if it is the better-than-standard approach but less accurate if it is the standard approach, etc.

Finally, it is possible that any relationships that exist between strategy appropriateness and strategy accuracy vary across mathematical domains. Procedural flexibility has been frequently studied in linear equation solving (e.g., Star & Seifert, 2006; Star & Rittle-Johnson, 2008; Newton et al., 2020) but also in fraction operations (Newton, 2008), calculus (Maciejewski & Star, 2016), integer operations (e.g., Torbeyns & Verschaffel, 2016), and other domains. It is unclear whether the substantial differences in these domains (including the age at which learners typically encounter these domains, as well as mathematical differences in the domains) might influence the relationships between strategy accuracy and strategy appropriateness.

With respect to mathematical domain differences in the relationship between strategy appropriateness and accuracy, of particular interest here is whether such differences exist between arithmetic problems and algebraic problems. Stu-

dents employ myriad strategies for computing in arithmetic, including both formal algorithms and informal strategies. With respect to the latter, the richness and variety of children's informal strategies for solving arithmetic problems has been well-documented in the literature (e.g., Shrager & Siegler, 1998). For symbolic algebra problems such as equation solving, students tend to demonstrate a greater reliance on formal symbolic algorithms (e.g., Mayer, 1982). For the reasons suggested above, it may be the case that standard algorithms are more accurate for both arithmetic and algebraic problems. However, one might also predict that there will be differences between these domains. On the one hand, the prevalence of a wide variety of informal and innovative strategies for arithmetic problems – and the increasing instructional emphasis on these strategies over the past several decades, especially in the U.S. (e.g., Carroll, 1999) – might suggest that standard arithmetic algorithms are on average less accurate than other strategies. But on the other hand, because of the greater abstraction present in algebra problems and the challenges that students often experience when learning algebra (e.g., Chu et al., 2017; Payne & Squibb, 1990; Van Amerom, 2003), it may be the case that standard algorithms are more accurate than alternative strategies in this mathematical domain.

In sum, there is considerable potential nuance in the relationship between strategy appropriateness and strategy accuracy within the context of procedural flexibility. Although there has been an increase in interest in and research on procedural flexibility over the past decade, little is known about this relationship.

Increasing our knowledge base about the relationship between strategy appropriateness and strategy accuracy is important to the field for the following reasons. First, scholars, educators, and policymakers have focused on increasing students' ability to employ multiple strategies flexibly in problem solving; however, flexibility itself absent the dimension of accuracy in problem solving seems a less desirable outcome. Arguments in favor of flexibility as an instructional goal would be substantially advanced if there were clearer links between flexibility and accuracy, including relationships between strategy appropriateness and accuracy. Attending to the dimension of accuracy in efforts to advance flexibility in practice would better align with broader goals for improving students' mathematical proficiency and conceptual understanding.

Second, understanding more about the relationship between strategy appropriateness and accuracy also relates directly to debates in mathematics education about the role of standard algorithms in the curriculum (e.g., Ebby, 2005; Van den Heuvel-Panhuizen, 2010). Algorithms are powerful tools, but over-reliance instructionally on algorithms has been linked in some countries to rote and inflexible knowledge. Under what conditions do standard algorithms offer benefits to students in terms of problem-solving accuracy? Under what conditions do better strategies not only indicate awareness of mathematical structure and deeper conceptual understanding but also yield more accurate results? These questions are core to conversations about what we teach in math and why we teach it.

Current Study

The relationship between strategy appropriateness and strategy accuracy is the focus of the present study. We explore this relationship in the context of algebra equation solving problems and in arithmetic problems, as well as in a task that prompted students to solve problems in different ways. While prior literature has demonstrated the variability in individual students' strategy choices on the same items and across occasions (e.g., Siegler, 1998), we were interested in examining how accuracy varies between standard and above-standard approaches on a task in which students are explicitly asked to show variable strategy use. We ask the following research questions in this study. First (R1), are students more accurate when using standard approaches or better-than-standard approaches? Second (R2), is this relationship between accuracy and strategy appropriateness influenced by whether a problem is being solved for the first time or being re-solved? Third (R3) are students more accurate when solving problems that are algebraic or arithmetic? We did not begin the study with *a priori* hypotheses about the answers to the research questions given the considerable nuances and uncertainties described above.

Method

Participants

A convenience sample of 450 high school students from 19 math classes in a single large high school in the Southeastern region of the United States participated in this study. We reduced the sample to the 449 students who completed the assessment. An additional 36 students who were missing demographic data were also excluded, leaving $N = 413$ students in the full sample. Data were collected in January 2020. The girl-boy ratio was almost evenly distributed, with 47.2% girls and 52.5% boys (see Table 1). About 53% of participants were in 9th grade, 26% were in 10th grade, and 21% were in 11th or 12th grade at the time of the study. Students were taking Algebra 1, Algebra 2, Geometry or AP Statistics courses. Students' self-reported grades in math ranged from 41.2% earning A's, 32.7% earning B's, 17.7% earning C's, and 8.5% earning a D or lower. The majority of students, 86.9%, were between the ages of 14 and 16 years old.

Table 1

Descriptive Statistics for Students in the Full Sample

Student Demographics	%
Gender	
Girls	47.2%
Boys	52.5%
Non-Binary	< 1.0%
Grade in School	
9th	53.3%
10th	26.2%
11th	13.6%
12th	7.0%
Self-reported Math Grade	
A	41.2%
B	32.7%
C	17.7%
D	6.5%
F	1.9%
Age	
13	0.2%
14	27.6%
15	39.5%
16	19.9%
17	9.9%
18	2.9%

Note. $N = 413$.

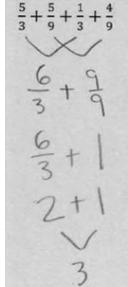
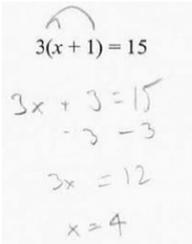
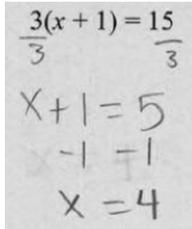
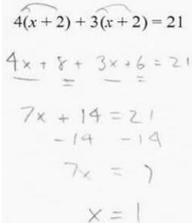
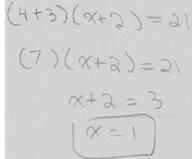
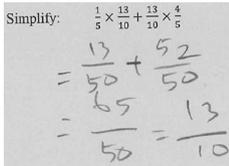
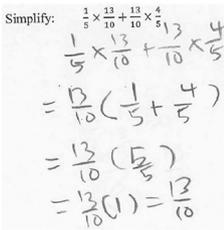
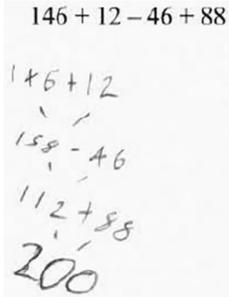
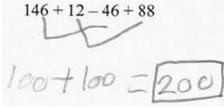
Measures

Participants completed a two-part assessment. In Part 1, students were prompted to complete five problems (see Table 2), each of which could be approached with standard and better-than-standard approaches. For example, in Problem 5, a standard approach is to add the numbers from left to right. Alternatively, a better-than-standard approach would be to commute terms and add $(12 + 88)$ and $(146 - 46)$, using 100 as a reference point to simplify the operations. The instructions prompted students not to proceed to the next section until instructed to do so. Then in Part 2, the assessment presented students with the same problems from Part 1 and instructed them to complete each problem using a different method than the one they used before. Students were instructed not to look back at their work in Part 1 while completing the same set of problems in Part 2. Students completed the 5 problems in each of the two parts of the

assessment, yielding 4,130 student-problems (resulting from 413 students completing each of 5 questions twice) in the data set. The unit of analysis for this study was student-problems.

Table 2

Problems Attempted in Parts 1 and 2 of the Assessment and Examples of Standard and Better-Than-Standard Solution Types

Accuracy	Problem	Standard Approach	Better-than-Standard Approach
1	Simplify: $\frac{5}{3} + \frac{5}{9} + \frac{1}{3} + \frac{4}{9}$	Simplify: $\frac{5}{3} + \frac{5}{9} + \frac{1}{3} + \frac{4}{9}$ 	
2	Solve: $3(x+1) = 15$		
3	Solve: $4(x+2) + 3(x+2) = 21$		
4	Simplify: $\frac{1}{5} \times \frac{13}{10} + \frac{13}{10} \times \frac{4}{5}$	Simplify: $\frac{1}{5} \times \frac{13}{10} + \frac{13}{10} \times \frac{4}{5}$ 	Simplify: $\frac{1}{5} \times \frac{13}{10} + \frac{13}{10} \times \frac{4}{5}$ 
5	Simplify: $146 + 12 - 46 + 88$	$146 + 12 - 46 + 88$ 	$146 + 12 - 46 + 88$ 

Problems 2 and 3 represent algebra problems, and Problems 1, 4 and 5 represent arithmetic problems. Our measure of the algebra domain consisted of both replications of Items 2 and 3 across the two assessment parts, totaling 4 items ($\alpha =$

.60). Similarly, our measure of the arithmetic domain consisted of both replications of Items 1, 4 and 5 across the two assessment parts, totaling 6 items ($\alpha = .60$).

Coding

We coded student-problems for both accuracy and type of strategy; of particular interest here is the distinction between standard and better-than-standard strategies. Two coders independently coded all strategies for strategy type and accuracy and subsequently resolved all disagreements. We elaborate on these types of strategies below and provide student examples in Table 2.

The standard approach for Problem 1 involved putting all the fractions in the same form over the common denominator of 9 to get $15/9$, $5/9$, $3/9$, and $4/9$, adding the numerators to get $27/9$, and finally simplifying to get 3 (see Table 2). For Problem 2, the standard approach involved distributing the 3, subtracting three from both sides, and at the end dividing both sides by 3 to obtain $x = 4$. Problem 3 follows a similar approach to problem Number 2. Using the standard approach, students first distribute the coefficients 4 and 3 respectively to get the expression $4x + 8 + 3x + 6$ on the left side before combining like terms, subtracting 14 from both sides, and finally dividing both sides by 7 to obtain $x = 1$. For Problem 4, the standard approach is to first multiply both sets of fractions in line with the order of operations, adding $13/50$ with $52/50$ to get $65/50$, and simplifying to obtain $13/10$ or the mixed number 1 and $3/10$. Finally, the standard approach for the last problem involves computing each operation in the expression from left to right to arrive at 200.

Student-problems were coded as better-than-standard if their approach demonstrated more elegance and innovation than the standard approach, based on similar determinations in prior studies (e.g., Star & Seifert, 2006; Star et al., 2022). For example, for Question 2, if a student divided by 3 as a first step, this was considered a better-than-standard strategy because the strategy takes advantage of the structural features of the problem (15 is evenly divisible by 3) and can be solved in fewer steps. In Question 3, as another example, if a student noticed that $4(x + 2)$ and $3(x + 2)$ were like terms and combined them as a first step to create $7(x + 2)$ on the left side of the equation, this was coded as better-than-standard. See Table 2 for additional examples.

Because our research questions were concerned with differences in accuracy between standard and better-than-standard approaches, we restricted the analysis sample to student-problems that used either of these two types of strategies across both Parts 1 and 2. Given the well-established relationship between worse-than-standard strategies and inaccuracy, we were only interested in examining problem solving accuracy between standard or better strategies (e.g., Star et al., 2022). Student-problems that were not included in the subsequent analysis included those that were incomplete, left blank, or showed a strategy that was judged by coders to be worse than the standard strategy on at least one part of the assessment. An example of a worse-than-standard strategy is using tallying to simplify an expression containing large integers (e.g., counting up 12 tallies or “tick” marks from 146 to add the expression $146 + 12$). Excluding the 2,046 student-problems (or 1,023 student-questions) that were coded as having used neither the standard nor better-than-standard approaches across either of the two assessment parts, we retain $n = 2,084$ student-problems (from 1,042 student-questions and $n = 377$ students) in the analysis sample. Of the 1,319 individual student-problems coded for demonstrating a below-standard strategy, 91.2% were marked incorrect, underscoring our need for excluding them from our analysis sample. See Table 3 for a description of the patterns of strategy use across Parts 1 and 2 for the full sample. We found no significant differences in student demographics between the full and analysis samples using chi-square difference tests.

Analysis

To begin investigating differences in accuracy between standard and better-than-standard approaches (RQ1), we present descriptive statistics for the accuracy rate by strategy use in student-problems. To begin to answer our second research question (R2) concerning how the relationship between strategy selection and accuracy might differ depending on assessment part, we present the accuracy rate for each strategy within Parts 1 and 2 of the assessment separately. We also show patterns of strategy use and accuracy across the problems in the first and second attempt to contextualize student-problems in the two-part assessment task. We present similar descriptives for the accuracy rate by problem type

to begin answering our third research question (RQ3) concerning differences in problem solving success on algebra problems compared to arithmetic ones.

Table 3

Patterns of Strategy Use Across Parts 1 and 2 in the Full Sample

Part 1	Part 2	Total (%)
Below Standard	Below Standard	296 (14.3)
Below Standard	Standard	47 (2.3)
Below Standard	Above Standard	53 (2.6)
Standard	Below Standard	520 (25.2)
Standard	Standard	271 (13.1)
Standard	Above Standard	487 (23.6)
Above Standard	Below Standard	107 (5.2)
Above Standard	Standard	160 (7.7)
Above Standard	Above Standard	124 (6.0)
		2,065

Note. $n = 2,065$ student-questions for $N = 413$ students. Strategies in bold indicate student-questions retained in the analysis sample ($n = 1,042$ student-question for $n = 377$ students).

We then fit a multi-level logistic regression model, with the dichotomous outcome variable corresponding to whether students' answer to each problem was correct or incorrect (see Equation 1). We created a three-level mixed-effects model with random intercepts across students and classrooms to account for the nesting of student-problem i within student j within classroom k . This modeling of within and between student, and within and between classroom, differences in problem-solving success enabled us to account for the homogeneity of errors within each group (Raudenbush & Bryk, 2002). The equation for our main model is represented as:

$$\text{logit}(\pi_{ijk}) = \beta_0 + \beta_1 \text{algebra}_{ijk} + \beta_2 \text{abovestand}_{ijk} + \beta_3 \text{part}_{ijk} + \beta_4 \text{abovestand}^* \text{part}_{ijk} + \epsilon_{ijk} + u_{0jk} + u_{00k} \quad (1)$$

where $\pi_{ijk} = Pr(\text{correct}_{ijk} = 1)$; ϵ_{ijk} , u_{0jk} , and u_{00k} are the Level 1, 2, and 3 variance components, respectively; $\epsilon_{ijk} = N(0, \sigma^2)$; $j = 1, \dots, 377$ students with $i = 2, \dots, 10$ student-problems in each student, and $k = 1, \dots, 20$ classrooms with $i = 6, \dots, 186$ student-problems in each classroom. To ensure the appropriateness of a three-level mixed-effects model, we first fit a null (unconditional) model to test whether our outcome of interest, accuracy of student-problem, varied across students and classrooms. We calculated the student-level and classroom-level intraclass correlation coefficients (ICC) to determine the proportion of variance accounted for between students and between classrooms (Raudenbush & Bryk, 2002). The ICC at the classroom level indicated that 6.44% of the chance of answering correctly was explained by between-classroom differences (and 93.56% was explained by within-classroom differences); the ICC at the student level showed that 29.14% of the chance of answering correctly was explained by between-student differences (and 70.86% was explained by within-student differences). We thus proceeded with a three-level model.

To answer our research questions about how accuracy may relate to strategy selection (RQ1) and problem type (RQ3), we included three dichotomous level one variables in the model: the strategy employed (standard or better-than-standard), assessment part (Part 1 or Part 2), and problem type (arithmetic or algebra). To fully answer RQ2, we modeled the interaction between the strategy used and assessment part to determine if the effect of strategy on accuracy depended upon whether the student was on Part 1 or Part 2 of the exam, in which students were asked to re-solve the same problem a second time using a different strategy from the one employed in Part 1. Finally, we conducted post-hoc tests to test for differences in the likelihood of an accurate response. All computing was completed using Stata version 17.0.

Results

Table 4 presents the frequency counts and percentages correct by strategy for student-problems in the analysis sample. Reliance on the standard algorithm was common across all problems, with 57.1% of all student-problems using this strategy compared to 42.9% employing better-than-standard ones. Of the 1,189 student-problems that used the standard approach, 80.7% were completed accurately, compared to 74.2% correct for the 895 student-problems that used better-than-standard approaches.

Table 4

Frequency Counts and Percentages Correct by Strategy, Problem Type, and Assessment Part for the Analysis Sample

Accuracy	Strategy Type						Problem Type		Assessment Part	
	Analysis Sample (N = 2,084)		Part 1 Only (n = 1,042)		Part 2 Only (n = 1,042)		Analysis Sample (N = 2,084)		Analysis Sample (N = 2,084)	
	Stand.	Above- Stand.	Stand.	Above- Stand.	Stand.	Above- Stand.	Arithmetic Problems	Algebra Problems	Part 1	Part 2
Incorrect	230 (19.3%)	231 (25.8%)	115 (15.2%)	86 (30.3%)	115 (26.7%)	145 (23.7%)	275 (24.9%)	186 (19.0%)	201 (19.3%)	260 (25.0%)
Correct	959 (80.7%)	664 (74.2%)	643 (84.8%)	198 (69.7%)	316 (73.3%)	466 (76.3%)	831 (75.1%)	792 (81.0%)	841 (80.7%)	782 (75.1%)
Total	1,189	895	758	284	431	611	1,106	978	1,042	1,042

Note. N = 2,084. Frequency counts and percentages correct by strategy are also shown separately for assessment Part 1 (n = 1,042) and Part 2 (n = 1,042). Column percentages are shown in parentheses.

However, we observed notable differences in accuracy by assessment part. In Part 1, the accuracy rate for student-problems that used standard approaches was 84.8%, compared to 69.7% for better-than-standard approaches. But in Part 2, the difference between the accuracy obtained between both approaches reverses, with standard and better-than-standard strategies demonstrating accuracy rates of 73.3% and 76.3%, respectively.

With regard to differences in accuracy by assessment part alone, we found that 80.7% of student-problems in Part 1 were completed correctly, compared to 75.1% in Part 2. Further, when it came to problem type, 75.1% of arithmetic student-problems were completed correctly, compared to 81.0% of algebraic ones. Students appeared to be more successful on algebra problems compared to arithmetic ones.

Table 5 describes patterns of accuracy and strategy across Parts 1 and 2 for all student-questions (n = 1,042). The majority of student-questions (46.7%) showed the student using a standard approach in Part 1 followed by an above-standard approach in Part 2, and the proportion of these student-questions answered correctly in both parts of the exam was 74.7%. 15.4% of student-questions showed the opposite pattern, with the student starting with the above-standard approach in Part 1 followed by a standard strategy in Part 2, and the proportion of these student-question answered correctly both times was comparable at 73.1%.

Table 5

Frequency Counts for Patterns of Accuracy and Strategy Across Parts 1 and 2

Strategy Across Parts 1 and 2		Accuracy Across Parts 1 and 2				Total
Part 1	Part 2	Both Correct	Part 1 Correct Only	Part 2 Correct Only	Both Incorrect	
Standard	Standard	166 (61.3%)	40 (14.8%)	21 (7.7%)	44 (16.2%)	271 (26.0%)
Standard	Above Standard	364 (74.7%)	73 (15.0%)	27 (5.5%)	23 (4.7%)	487 (46.7%)
Above Standard	Above Standard	56 (45.2%)	10 (8.1%)	19 (15.3%)	39 (31.5%)	124 (11.9%)
Above Standard	Standard	117 (73.1%)	15 (9.4%)	12 (7.5%)	16 (10.0%)	160 (15.4%)
		703	138	79	122	1,042

Note. n = 1,042 Student-Questions. Row percentages are shown in parentheses. Column percentages are shown for the final column.

The results from our mixed-effects logistic regression analysis elaborate on these findings. We begin with our third research question concerning differences in accuracy by problem type, which factors into our discussion of the findings for our first two research questions. We found further support for the finding that responses on algebra problems were more accurate than responses for arithmetic problems, even when controlling for assessment part and strategy type. Algebra problems had an estimated 0.34-logit greater likelihood of a correct response compared to arithmetic problems, $\beta_1 = 0.34$, $z = 2.54$, $p = .011$; see Table 6. In the next sections, we describe the probability of a correct response by strategy type and assessment part for algebraic and arithmetic student-problems separately.

Table 6

Mixed-Effects Logistic Regression Results Showing Prediction of Accuracy in Log-Odd Units

Predictor	<i>b</i>	<i>SE</i>	95% CI	
			<i>LL</i>	<i>UL</i>
(Intercept)	1.89***	0.20	1.49	2.29
Algebra	0.34*	0.13	.08	.60
Above-Standard Approach	-0.90***	0.20	-1.30	-.51
Part 2 of Assessment	-0.70***	0.17	-1.03	-.36
Above-Standard Approach*Part 2	0.89**	0.26	.38	1.40

Note. $N = 2,084$ Student-Problems. Standard errors in parentheses.

* $p < .05$. ** $p < .01$. *** $p < .001$.

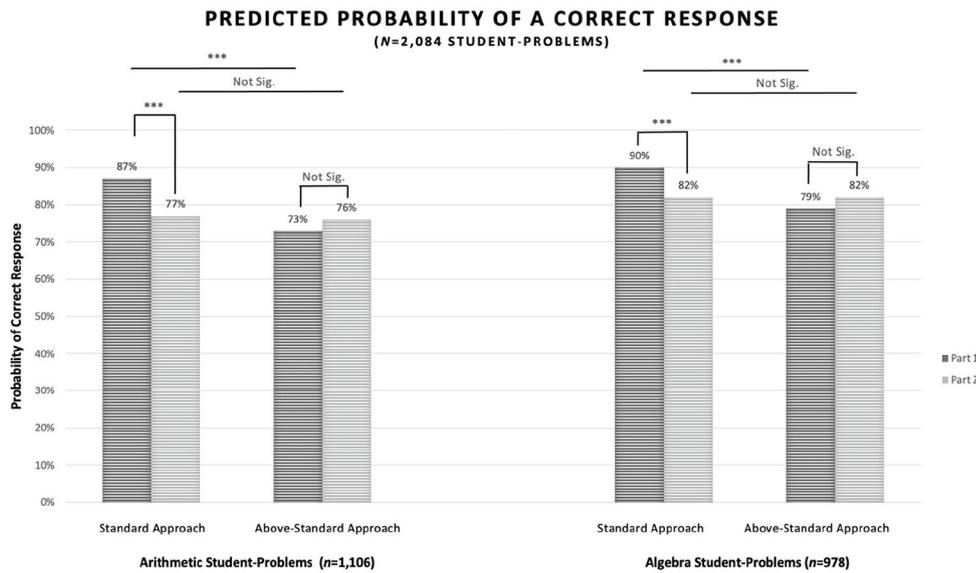
With regard to our first research question concerning differences in accuracy by strategy type and our second research question investigating how this relationship may differ by assessment part, we found further evidence in support of the interaction between strategy use and whether a student was completing a problem for the first or second time. In Part 1 of the assessment, the standard approach was related to a greater likelihood of an accurate response compared to the above-standard approach, $\beta_2 = -.90$, $z = -4.50$, $p < .001$ (see Table 6). For arithmetic student-problems in Part 1, the standard approach was associated with an estimated 87% chance of a correct response, compared to only a 73% chance with above-standard approaches. For algebra student-problems in Part 1, standard and above-standard approaches were associated with an estimated 90% and 79% probability of success, respectively. See Figure 1 for the estimated probabilities of a correct response by strategy type for arithmetic and algebra problems separately.

However, this relationship between strategy use and accuracy changes in Part 2, as evidenced by the significant interaction term between strategy and part, $\beta_4 = 0.89$, $z = 3.40$, $p = .001$ (see Table 6). When students solved the same problems a second time in Part 2 of the assessment, the differences in accuracy by strategy type disappear. For arithmetic problems in Part 2, standard and above-standard approaches were associated with an estimated 77% and 76% likelihood of a correct response, respectively; for algebra problems in Part 2, both strategy types were associated with an estimated 82% chance of being answered correctly (see Figure 1). For arithmetic and algebra problems alike, in Part 2 of the assessment we found no significant differences in accuracy between standard and better-than-standard approaches, $X^2(1, N = 2,084) = 0.01$, $p = .9221$.

Examining the effectiveness of each strategy type across the two attempts, the standard approach was significantly more likely to yield a correct response in Part 1 compared to Part 2, but this is not the case for the above-standard approach, which demonstrated equal chances of success in both Parts 1 and 2. The standard approach had an estimated 10% greater chance of success in Part 1 (87%) compared to Part 2 (77%) for arithmetic problems and an estimated 8% greater chance of success in Part 1 (90%) compared to Part 2 (82%) for algebra problems, $\beta_3 = -.70$, $z = -4.11$, $p < .001$ (see Table 6). For above-standard approaches, we found no differences in the chance of obtaining a correct response across Parts 1 and 2, $X^2(1, N = 2,084) = 1.02$, $p = .3123$. Above-standard approaches had an estimated 73% chance of success in Part 1 and 76% chance of success in Part 2 for arithmetic problems, and an estimated 79% chance of success in Part 1 and 82% chance in Part 2 for algebra problems. See Figure 1 for the estimated probabilities of a correct response by strategy, assessment part, and problem domain.

Figure 1

Proportion of Student-Problems Correct by Strategy, Assessment Part, and Problem Type



Note. Test results for differences in the likelihood of a correct response between strategies within the *same part* of the assessment are shown in the bars above. Test results for differences in the likelihood of a correct response between the two assessment parts within the *same strategy* are shown in brackets.

* $p < .05$. ** $p < .01$. *** $p < .001$.

Discussion

In the present study, we investigated differences in the accuracy achieved with standard and better-than-standard strategies and the extent to which accuracy differed by whether a student was solving for the first or second time and whether the problem was arithmetic or algebraic. We found that there is a relationship between strategy use and accuracy, and that the relationship depends on whether students are doing a problem for the first or second time as well as on the problem domain. The standard approach was related to greater success in problem solving only when a student was solving a problem for the first time; when asked to solve the same problems a second time using a different strategy, the standard approach was no more accurate than better-than-standard approaches. Further, examining the patterns of strategy use across the problems in the first and second attempt showed a majority of student-problems beginning with the standard approach in Part 1 followed by an above-standard approach in Part 2, with the proportion answered correctly both times being 74.7%. Students' flexible strategy use appears to show the dominance of the standard approach as the primary strategy of choice. A small proportion of student-questions began with the above-standard approach as the primary choice of strategy in Part 1 followed by a standard strategy in Part 2, with the proportion answered correctly both times comparable at 73.1%.

For students' primary strategy selection in Part 1, the standard approach may be associated with a higher accuracy rate than the better-than-standard strategy for reasons that may seem intuitive: this approach is the more common, reliable, and routine way of solving a problem. The success of the standard approach here may be attributable to the "freed resources account" for student attention on problems, which posits that highly routinized and well-practiced strategies require fewer cognitive resources from students, "freeing up" their ability attend to the key features and relationships in the problem (Kotovsky et al., 1985; McNeil & Alibali, 2004; Shiffrin & Schneider, 1977; Shrager & Siegler, 1998). Students who used the better-than-standard strategy in Part 1 may have had fewer cognitive resources at their disposal given the nonroutine nature of these approaches, making it more difficult to attend to the mechanics of

solving and simplifying. This “freed resources account” may have made the problem easier to solve using the standard approach— and more likely to lead to an accurate solution— compared to better-than-standard approaches in Part 1.

However, when students were asked to go beyond their primary strategy of choice, as our assessment prompted them to do in Part 2, both strategy types were equally related to accuracy. To add to this, the standard approach was significantly more successful in Part 1 compared to Part 2, but better-than-standard approaches were equally successfully across the two exam parts. It could be the case that the application of above-standard approaches is a robust indicator of greater flexibility and conceptual understanding, regardless of whether this type of approach is a student’s primary or secondary strategy choice. Recognizing the structural features in a problem is related to flexibility and conceptual knowledge (e.g., Schneider et al., 2011), and this may explain the stability of this approach’s relation to accuracy irrespective of which attempt a student was on. It could also be the case that students using better-than-standard strategies in Part 2 benefitted from having previously solved using the standard method, improving their accuracy the second time around. Further, students using the standard strategy the second time may have had less practice with this approach, as it was not their primary choice of strategy, and this lack of routinization could have contributed to the reduction in accuracy we found. When students do not have a well-routinized and practiced strategy readily available to them, they need to attend to the specific features of the problem more carefully and encode the information in the problem to devise their approach (McNeil & Alibali, 2004). This on-the-spot encoding could have taken up more working memory, increasing the likelihood for error. It could also be the case that students only familiar with standard approaches tried to apply the same standard procedures a second time but were less successful. This is because highly-practiced and internalized strategies— such as the standard ones— dictate a student’s encoding when problem-solving, and this top-down approach may cause them to only attend to and encode the features necessary for executing their specific strategy (McNeil & Alibali, 2004). This relationship between strategy use and encoding may have contributed to the reduction in accuracy we found in Part 2 with the standard approach compared to the same approach the first time.

Our results point to the value in cultivating flexibility in problem solving for learners. The vast majority of students in our sample who correctly solved a problem twice used some combination of standard and above-standard approaches across the two parts (see Table 5). This is not surprising, given that procedural flexibility is indicative of greater conceptual and procedural knowledge in mathematics (Durkin et al., 2021). Demonstrating successful problem solving using multiple strategies is a more robust indicator of student learning and comprehension than successful problem solving using one strategy alone. Understanding each strategy’s likelihood of success on a task that promotes flexible strategy use helps us get a clearer picture of students’ procedural flexibility.

The Relationship Between Problem Domain and Accuracy

While our findings more generally indicated that the success of certain strategies depends on whether a student is solving a problem for the first or second time, we observed notable differences in this relationship depending on the problem domain. We found significant differences in accuracy across algebraic and arithmetic student-problems, even controlling for assessment part and strategy type. It is possible this difference in accuracy is due to structural differences between the two problem types. The multi-step equations found in algebra domains may reduce the cognitive burdens of problem solving, given their predictable and less varied nature compared to arithmetic problems, with common patterns of distributing coefficients, combining like terms, and isolating variables. It seems reasonable for students to have common templates for solving, given the common format and structural features found in such equations.

It could also be the case that equations that are “flexible-eligible” (Hästö et al., 2019) may offer more entry points for solving. For arithmetic problems, particularly those containing fraction operations as seen in our assessment, students may have limited facility given the way fractions are taught in many U.S. schools (e.g., Harvey, 2012; Lamon, 2007). Another reason for the greater accuracy found in algebra problems could simply be recency effects: temporal distance from curricula emphasizing the fraction and integer operations more often seen in primary school mathematics may hinder students’ success when it comes to these same types of problems later in high school, where the mathematics curricular program is heavily focused on algebra.

Limitations and Future Directions

A limitation of the present study is the use of student-problems as the unit of analysis. Future studies could look across assessment questions for the same student and examine each student's strategy selection and accuracy conditionally on how the student solved a problem for the first time, including those that used a below-standard strategy. Given that we excluded 2,046 student-problems (or 1,023 student-questions) that showed a below-standard strategy in at least one part of the assessment, our study does not generalize to flexible problem-solving situations in which a student uses a standard or better strategy in combination with a below-standard strategy; our results can only speak to the relationship between flexible strategy use and accuracy for students using standard or better strategies. We recommend future work that examines the relationship between below-standard approaches, procedural flexibility, and conceptual understanding.

Future work investigating flexible strategy use might also adapt the choice/no-choice method, which compares problem solving latencies or times-to-solution between a method of choice and a prescribed method within individual students (Siegler & Lemaire, 1997). This technique could help to illuminate patterns in the relationship between accuracy and flexibility at the student level. The choice/ no-choice method has been widely applied in the study of flexible strategy use on a broad array of task types with a broad range of latencies. However, this method works best for mental computation or tasks with limited use of external aids, such as calculators and pencil and paper on short multiplication tasks (Siegler & Lemaire, 1997). Given that the hand-written symbolic manipulation seen in algebraic problem-solving increases latencies by up to a factor of 10 compared to prior studies using this method, the choice/ no-choice method would likely need to be adapted for the study of flexible strategy use in algebraic problem solving. Another limitation of the choice/ no-choice method is that researchers must explicitly prescribe a specific strategy to participants in the no-choice condition (Luwel et al., 2009). This is more easily done in, for example, simple arithmetic tasks by telling participants to round to the nearest tens or hundreds place. In algebra problem-solving, however, naming the strategy for the student without inadvertently aiding a student through problem solving may be difficult. Students are unlikely to know what the "standard approach" or the "above-standard approach" means, and to accurately prescribe this technique to students by describing the steps to them might undermine the point of investigating students' independent flexible strategy use. We recommend future methodological exploration of how to adapt the choice/ no-choice method to longer, hand-written algebraic and arithmetic problem-solving tasks such as the ones used in the current study.

In addition, further work could be done to explore more of the variation in strategy selection and accuracy within each problem type. For example, the two algebra equations we used, while typical in secondary mathematics and algebra curricula, do not capture the full array of algebraic problem features students encounter. Similarly, the arithmetic problems in our assessment are primarily concerned with fraction and integer operations. Our results may have been influenced more by the specifics of these problem features rather than arithmetic problems more generally, limiting the generalizability of our findings to these problem domains. Future studies exploring the relationship between accuracy and strategy choice in different problem domains may wish to better account for general characteristics of problems in the mathematical domains of interest as well as to increase the number of items and thus reliability. Similar to problem domain, future studies may wish to investigate the relationship between strategy selection and accuracy on word problems in mathematics. Prior work has shown that even expert mathematicians struggle to apply simple arithmetic procedures to word problems when the problem presents semantic content that is incongruent with the arithmetic solving procedure (Gros et al., 2019). Extending our research questions to word problem situations invoking flexible strategy use and application via problem statement would provide meaningful nuance to our understanding of the phenomenon, especially in this Common Core era of open-ended tasks and word problems in mathematics.

Another limitation of the current study is the limited age and grade range of students in the sample (86.9% of students in the sample were between the ages of 14 and 16 years old, and 79.5% were in grades 9 or 10), precluding us from examining the effect of student age, grade, and by proxy grade-based curriculum, on problem-solving success. We recommend a thorough examination of situational variables related to the learner, taking into account student characteristics such as age, grade, math placement, and math performance as they relate to procedural flexibility. Related to situational variables, our sample comes from one specific high school, limiting the generalizability of our findings— future studies on the relationship between flexible strategy use and accuracy may wish to vary the school site

and examine contextual factors related to the phenomenon. Finally, more qualitative research on students' encoding of flexible-eligible problems and how a student decides which strategy to apply is needed to better understand students' rationale for which strategies they employ. For example, secondary school students in Spain tend to prefer standard algorithms and approaches (Jiang et al., 2022).

Conclusion

Our findings have implications for mathematics educators seeking to promote procedural flexibility in their classrooms. There is potential value in using this kind of task (in which students are prompted to re-solve a previously completed problem) for two reasons. First, our results are consistent with prior calls for the inclusion of this type of task, both as a student learning task as well as an assessment task (e.g., Blöte et al., 2001; Star & Seifert, 2006). Being asked to re-solve a problem prompts students to try to generate a different strategy, which can have the effect of building knowledge of multiple strategies and thus potentially flexibility. But second, the presence of multiple strategies, as promoted by this task, affords teachers the potential opportunity to engage students in thinking and discussion around some of the nuanced issues related to flexibility. These include which strategy students feel that they can more accurately and reliably use, as well as which strategy is better (and what 'better' means). Engaging students in such metacognition of strategy appropriateness may deepen student mathematical knowledge and flexibility (Heirdsfield & Cooper, 2002). Further, a task of this nature may prompt learners to recognize and take advantage of the structural features in a problem that they may not have noticed when solving the first time, and this may help to develop their conceptual understanding and flexibility (Schneider et al., 2011). Our analysis contributes important nuance to our understanding of procedural flexibility with respect to the relationship between the use of standard and better-than-standard strategies and the accuracy of these strategies, adding complexity to debates about which strategies are better for problem solving.

Funding: The authors have no funding to report.

Acknowledgments: The authors have no additional (i.e., non-financial) support to report.

Competing Interests: The authors have declared that no competing interests exist.

References

- 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), 627–638. <https://doi.org/10.1037/0022-0663.93.3.627>
- Buchbinder, O., Chazan, D., & Fleming, E. (2015). Insights into the school mathematics tradition from solving linear equations. *For the Learning of Mathematics*, *35*(2), 2–8. <https://flm-journal.org/Articles/36150A50EA49573571336BD71678E5.pdf>
- Carroll, W. M. (1999). Invented computational procedures of students in a standards-based curriculum. *The Journal of Mathematical Behavior*, *18*(2), 111–121. [https://doi.org/10.1016/S0732-3123\(99\)00024-3](https://doi.org/10.1016/S0732-3123(99)00024-3)
- Chu, J., Rittle-Johnson, B., & Fyfe, E. R. (2017). Diagrams benefit symbolic problem-solving. *The British Journal of Educational Psychology*, *87*(2), 273–287. <https://doi.org/10.1111/bjep.12149>
- Durkin, K., Rittle-Johnson, B., Star, J. R., & Loehr, A. (2021). Comparing and discussing multiple strategies: An approach to improving algebra instruction. *Journal of Experimental Education*. Advance online publication. <https://doi.org/10.1080/00220973.2021.1903377>
- Ebby, C. B. (2005). The powers and pitfalls of algorithmic knowledge: A case study. *The Journal of Mathematical Behavior*, *24*(1), 73–87. <https://doi.org/10.1016/j.jmathb.2004.12.002>
- Gros, H., Sander, E., & Thibaut, J. P. (2019). When masters of abstraction run into a concrete wall: Experts failing arithmetic word problems. *Psychonomic Bulletin & Review*, *26*(5), 1738–1746. <https://doi.org/10.3758/s13423-019-01628-3>
- Hardy, G. H. (1940). *A mathematician's apology*. Cambridge University Press.
- Harvey, R. (2012). Stretching student teachers' understanding of fractions. *Mathematics Education Research Journal*, *24*, 493–511. <https://doi.org/10.1007/s13394-012-0050-7>

- 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. <https://doi.org/10.30935/scimath/9530>
- Heirdsfield, A. M., & Cooper, T. J. (2002). Flexibility and inflexibility in accurate mental addition and subtraction: Two case studies. *The Journal of Mathematical Behavior*, 21(1), 57–74. [https://doi.org/10.1016/S0732-3123\(02\)00103-7](https://doi.org/10.1016/S0732-3123(02)00103-7)
- Institute of Educational Sciences. (2012). *Improving mathematical problem solving in Grades 4 to 8: A practice guide* (NCEE 2012-4055). National Center for Education Evaluation and Regional Assistance, U.S. Department of Education.
- Jiang, R., Star, J. R., Hästö, P., Li, L., Liu, R., Tuomela, D., Joglar, N., Palkki, R., Abánades, M., & Pejlar, J. (2022). Which one is the “best”: A cross-national comparative study of students’ strategy evaluation in equation solving. *International Journal of Science and Mathematics Education*. Advance online publication. <https://doi.org/10.1007/s10763-022-10282-6>
- Kilpatrick, J., Swafford, J. O., & Findell, B. (Eds.). (2001). *Adding it up: Helping children learn mathematics*. National Academy Press.
- Kotovsky, K., Hayes, J. R., & Simon, H. A. (1985). Why are some problems hard? Evidence from Tower of Hanoi. *Cognitive Psychology*, 17(2), 248–294. [https://doi.org/10.1016/0010-0285\(85\)90009-X](https://doi.org/10.1016/0010-0285(85)90009-X)
- Lamon, S. J. (2007). Rational numbers and proportional reasoning. In F. K. Lester (Ed.), *Second handbook of research on mathematics teaching and learning* (pp. 629–667). Information Age.
- Luwel, K., Onghena, P., Torbeyns, J., Schillemans, V., & Verschaffel, L. (2009). Strengths and weaknesses of the choice/no-choice method in research on strategy use. *European Psychologist*, 14(4), 351–362. <https://doi.org/10.1027/1016-9040.14.4.351>
- Maciejewski, W., & Star, J. R. (2016). Developing flexible procedural knowledge in undergraduate calculus. *Research in Mathematics Education*, 18(3), 299–316. <https://doi.org/10.1080/14794802.2016.1148626>
- Mayer, R. E. (1982). Different problem-solving strategies for algebra word and equation problems. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 8(5), 448–462. <https://doi.org/10.1037/0278-7393.8.5.448>
- McNeil, N. M., & Alibali, M. W. (2004). You’ll see what you mean: Students encode equations based on their knowledge of arithmetic. *Cognitive Science*, 28(3), 451–466. https://doi.org/10.1207/s15516709cog2803_7
- National Council of Teachers of Mathematics. (2000). *Principles and standards for school mathematics*. NCTM.
- National Council of Teachers of Mathematics. (2014). *Procedural fluency in mathematics: A position of the National Council of Teachers of Mathematics*. Retrieved on January 18, 2021, from <https://www.nctm.org/Standards-and-Positions/Position-Statements/Procedural-Fluency-in-Mathematics/>
- Newton, K. J. (2008). An extensive analysis of preservice elementary teachers’ knowledge of fractions. *American Educational Research Journal*, 45(4), 1080–1110. <https://doi.org/10.3102/0002831208320851>
- Newton, K. J., Lange, K., & Booth, J. L. (2020). Mathematical flexibility: Aspects of a continuum and the role of prior knowledge. *Journal of Experimental Education*, 88(4), 503–515. <https://doi.org/10.1080/00220973.2019.1586629>
- Payne, S. J., & Squibb, H. R. (1990). Algebra mal-rules and cognitive accounts of error. *Cognitive Science*, 14, 445–481. https://doi.org/10.1207/s15516709cog1403_4
- Raudenbush, S. W., & Bryk, A. S. (2002). *Hierarchical linear models: Applications and data analysis methods* (2nd ed.). SAGE.
- Schneider, M., Rittle-Johnson, B., & Star, J. R. (2011). Relations among conceptual knowledge, procedural knowledge, and procedural flexibility in two samples differing in prior knowledge. *Developmental Psychology*, 47(6), 1525–1538. <https://doi.org/10.1037/a0024997>
- Shiffrin, R. D., & Schneider, W. (1977). Controlled and automatic human information processing, II: Perceptual learning, automatic attending, and a general theory. *Psychological Review*, 84(2), 127–190. <https://doi.org/10.1037/0033-295X.84.2.127>
- Shrager, J., & Siegler, R. (1998). SCADS: A model of children’s strategy choices and strategy discovery. *Psychological Science*, 9, 405–410. <https://doi.org/10.1111/1467-9280.00076>
- Siegler, R. S. (1998). *Emerging minds: The process of change in children's thinking*. Oxford University Press.
- Siegler, R. S., & Lemaire, P. (1997). Older and younger adults’ strategy choices in multiplication: Testing predictions of ASCM using the choice/no-choice method. *Journal of Experimental Psychology: General*, 126(1), 71–92. <https://doi.org/10.1037/0096-3445.126.1.71>
- Star, J. R. (2005). Reconceptualizing procedural knowledge. *Journal for Research in Mathematics Education*, 36(5), 404–411.
- Star, J. R. (2007). Foregrounding procedural knowledge. *Journal for Research in Mathematics Education*, 38(2), 132–135.
- Star, J. R., & Rittle-Johnson, B. (2008). Flexibility in problem solving: The case of equation solving. *Learning and Instruction*, 18(6), 565–579. <https://doi.org/10.1016/j.learninstruc.2007.09.018>

- Star, J. R., & Seifert, C. (2006). The development of flexibility in equation solving. *Contemporary Educational Psychology*, 31(3), 280–300. <https://doi.org/10.1016/j.cedpsych.2005.08.001>
- Star, J. R., Tuomela, D., Joglar-Prieto, N., Hästö, P., Palkki, R., Abánades, M. A., Pejlare, J., Jiang, R. H., Li, L., & Liu, R.-D. (2022). Exploring students' procedural flexibility in three countries. *International Journal of STEM Education*, 9(1), Article 4. <https://doi.org/10.1186/s40594-021-00322-y>
- Torbeyns, J., & Verschaffel, L. (2013). Efficient and flexible strategy use on multi-digit sums: A choice/no-choice study. *Research in Mathematics Education*, 15(2), 129–140. <https://doi.org/10.1080/14794802.2013.797745>
- Torbeyns, J., & Verschaffel, L. (2016). Mental computation or standard algorithm? Children's strategy choices on multi-digit subtractions. *European Journal of Psychology of Education*, 31(2), 99–116. <https://doi.org/10.1007/s10212-015-0255-8>
- Van Amerom, B. A. (2003). Focusing on informal strategies when linking arithmetic to early algebra. *Educational Studies in Mathematics*, 54(1), 63–75. <http://www.jstor.org/stable/3483215><https://doi.org/10.1023/B:EDUC.0000005237.72281.bf>
- Van den Heuvel-Panhuizen, M. (2010). Reform under attack: Forty years of working on better mathematics education thrown on the scrapheap? No way! In L. Sparrow, B. Kissane, & C. Hurst (Eds.), *Shaping the future of mathematics education: Proceedings of the 33rd annual conference of the Mathematics Education Research Group of Australasia*. MERGA.
- Verschaffel, L., Luwel, K., Torbeyns, J., & Van Dooren, W. (2009). Conceptualizing, investigating, and enhancing adaptive expertise in elementary mathematics education. *European Journal of Psychology of Education*, 24(3), 335–359. <https://doi.org/10.1007/BF03174765>
- 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, Article 1368. <https://doi.org/10.3389/fpsyg.2017.01368>



Journal of Numerical Cognition (JNC) is an official journal of the Mathematical Cognition and Learning Society (MCLS).



leibniz-psychology.org

PsychOpen GOLD is a publishing service by Leibniz Institute for Psychology (ZPID), Germany.