# Understanding Students' Computational Fluency: Synechistically Using Test Scores and Interviews for a Richer Picture

Jessica F. Shumway & Kerry E. Jordan
*Utah State University*
jessica.shumway@usu.edu, kerry.jordan@usu.edu

The purpose of this study was to explore the variations in students' achievement on computational fluency assessments and variations in strategies for solving computation problems. Computational fluency was defined as a student's ability to compute flexibly, accurately, efficiently, and appropriately. Seven- and eight-year-old students completed four paper-pencil assessments and participated in four interviews about their strategies for solving computation problems. Researchers used a mixed methods design using test scores and interviews that measured six students' development of computational fluency at four time points over nine weeks. The results showed that there was variation in students' assessment scores and strategy use, and the quantitative and qualitative data converged or diverged at various measurement points. The divergent results highlighted a trend of improved strategy use in the interviews parallel to a decrease in test scores. Implications for mathematics teachers are discussed.

## Introduction

Erlwanger's (1973) classic mathematics education article, "Benny's conception of rules and answers in IPI mathematics," is a case study about a twelve-year-old student who appeared to be making better than average progress on basic math skills tests. However, through a series of discussions and interviews, Erlwanger uncovered Benny's lack of understanding hiding under his seeming mastery of computation skills. In the 1970s, Erlwanger's article not only opened a window to children's thinking and learning of mathematics, but also to the possibilities of research methods for better understanding children's mathematical thinking. Many of the pending research questions are not only about children's thinking nor only about "what works" in classrooms. There is a continuum between these two paradigms and linking the two can provide more complete understandings about how and why instructional practices influence children's learning of mathematics. Understanding variations of children's mathematical thinking and how they respond to instruction in classrooms can lead to effective instructional practices that meet the needs of students with wide-ranging mathematics readiness levels. Mixed methods in mathematics education research are being used to solve the complex, practical research problems that emerge in an applied discipline that grapples with *what* works and *how* mathematics learning results can be achieved based on what we know about children's mathematical thinking.

This article is about an examination of seven- and eight-year-old students' development of computational fluency. We begin the article with a discussion about computational fluency and how it is assessed. Then we discuss our study, explicitly laying out our reasons for using quantitative and qualitative data sources and mixing the analysis and results in order to gain a richer understanding of students' computational fluency. Finally, we conclude the article with limitations and educational implications.

## Computational Fluency

The issue of students knowing their basic facts was at the center of the 1990s "math wars" debates in the United States (Schoenfeld, 2004). For the most part, educators have moved beyond the math wars and agreed that teaching for both skill and understanding is essential, and that they can be learned together (Baroody, Feil, & Johnson, 2007; National Council of Teachers of Mathematics (NCTM), 2000). The term, *computational fluency*, is one construct in US schools that emphasizes the balance of skills and understanding. This term is meant to encompass both conceptual understanding and procedural fluency (Common Core State Standards Initiative (CCSSI), 2010; NCTM, 2000). Specifically, the Common Core State Standards for Mathematics (CCSSM) document delineates the expectation that students' computational fluency is defined as the "ability to compute flexibly, accurately, efficiently, and appropriately" (CCSSI, 2010, p. 6).

### *Assessing Computational Fluency*

Computational fluency has traditionally been assessed in US classrooms through timed paper-pencil assessments that measure accuracy and efficiency (e.g. Fuchs, Hamlett, & Powell, 2003). This method is also used in research (e.g. Göbel, Watson, Lervag, & Hulme, 2014). Rather than paper-pencil assessments, some studies use interview settings to assess students' computational fluency in terms of accuracy and efficiency. Number combinations are read to students, and students respond verbally with a solution (e.g. Jordan, Hanich, & Kaplan, 2003). In both cases, these studies involve hundreds of participants and are methodologically quantitative, hence, students' reasoning strategies and processes for coming to solutions are not studied.

Other studies operationalize computational fluency with measures for accuracy and proficiency while including analyses of students' reasoning strategies (e.g. Geary, 2011; Geary, Hoard, Byrd-Craven, & DeSoto, 2004; Jordan, Kaplan, Ramineni, & Lockuniak, 2009). And, other research has sought to examine how children's computational fluency and strategies predict performance on more complex arithmetic tasks (Vasilyeva, Laski, & Shen, 2015). These studies typically operationalize and assess *computational fluency* more in line with current classroom instruction goals that value both skill and understanding.

### *Assessing Computation Strategies*

Baroody (2006) described three phases students move through as they develop strategies for computational fluency: Phase 1 - Counting Strategies, Phase 2 - Reasoning Strategies, and Phase 3 - Mastery. Much of the debate in the 1990s dealt with how to get to Phase 3 - Mastery. Baroody, Bajwa, and Eiland (2009) explained two perspectives on attaining mastery. One perspective is the Passive Storage View, which views memorizing a basic number combination as a simple form of learning. From this view, Phase 3 - Mastery is achieved through memorization and drill practices. An alternative perspective, the Active Construction View, sees the memorization of a basic combination as a more complex form of learning that involves constructing a connected body of knowledge involving patterns, relations, algebraic rules, and automatic reasoning processes. Baroody and colleagues theorized that fluency with basic number combinations stems from students' number sense. Developing strategies based in number sense would occur during Phase 2 – Reasoning Strategies. This Active Construction View supports the notion of computational fluency as more than accuracy and efficiency, to one that includes strategies that are used flexibly and appropriately. From the Active Construction View, Phase 2 -

Reasoning Strategies is a critical learning phase for students on their path to computational fluency. We were interested in seven- and eight-year-old students' development of reasoning strategies from the perspective of the Active Construction View because the second-grade school year in the United States is a period of rapid development in their computational fluency.

### *Purpose of the Study*

The purpose of the study was to explore the variations in students' achievement on computational fluency assessments (i.e., fact fluency timed tests and untimed word problem tasks) and the variations in students' strategies for solving computation problems (i.e., counting, reasoning, or master) over the course of the study. To study the complex notion of computational fluency, defined as a student's ability to compute flexibly, accurately, efficiently, and appropriately, we collected both quantitative and qualitative data at four time points over nine weeks (Plano Clark et al., 2015). Seven- and eight-year-old students completed four paper-pencil assessments and participated in four interviews about their strategies for solving the problems on these assessments.

### *Research Questions*

We explored the following questions:

1. What are the variations within individual students' achievement scores on computational fluency assessments? (quantitative)
2. What are the variations within individual students' strategies for solving computation problems? (qualitative)
3. How do the results of the interviews converge, diverge, and/or augment the interpretation of the students' assessment scores? (mixed quantitative and qualitative)

## Methods

### *Methodological Framework*

This study is theoretically grounded in Baroody et al.'s (2009) Active Construction View. Within this perspective, conceptual and procedural knowledge are viewed within a continuum and intertwined (Baroody et al., 2007). Within the Active Construction View, paper-pencil assessments are limited in measuring students' computational fluency, and hence, this study combines paper-pencil assessments along with interviews to better understand students' construction of computational fluency. Rather than viewing these data sources as conflicting, we conceptualize this research process within a multidimensional continuum (Niglas, 2010).

Hence, this study uses convergent parallel mixed methods design; Figure 1 provides a diagram illustration of the design (Plano Clark et al., 2015). We collected quantitative data and qualitative data at four time points, analyzed the data separately, then merged the results of the two strands, looking for points when the interview data converged with, diverged from, and/or augmented the test score data (Creswell & Plano Clark, 2011). The purpose of utilizing this mixed methods design was to use both strands of data to develop a more complete understanding of students' computational fluency. The paper-pencil assessments were developed through a postpositivist lens in terms of operationalizing computational fluency into two main constructs: efficiency and accuracy with single-digit addition and subtraction problems. The semi-structured interview was developed through operationalizing computational fluency into additional constructs of flexibility in and appropriate use of reasoning strategies. While we recognize the influences of seemingly conflicting paradigms within different phases of the study, we think *synechistically* and favor viewing computational fluency in terms of continua rather than binaries (Johnson & Gray, 2010). We intentionally use the dialectics perspective to investigate both the divergences and convergences between data sets (Shannon-Baker, 2016). The dialectics perspective brings together two paradigms, allowing a broader understanding of students' computational fluency development, in terms of efficiency and accuracy, in conjunction with flexibility and appropriate use of reasoning strategies.

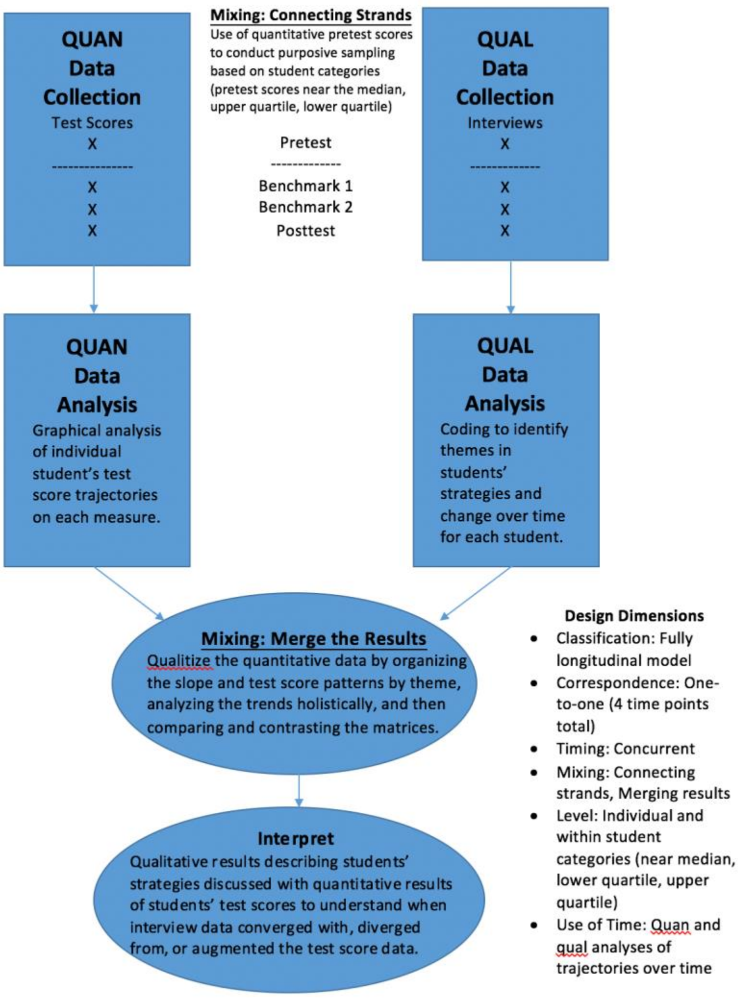


*Figure 1.* Flow of the methods in a mixed methods design.

*Note.* Design of the diagram was based on examples in Plano Clark et al. (2015).

### *Participants and Sampling Procedures*

The participants in this study were six second-grade students from one classroom in a public school located in the western United States. Students ranged in age from 7 years and 4 months to 8 years old at the outset of the study (November of their school year). Students' names have been replaced with pseudonyms.

In the initial phase of the study, all 21 students from one second-grade classroom were given the pretest assessment. A purposive range sample of six students was selected for individual interviews (Bamberger, Rugh, & Mabry, 2012). Researchers selected a sample of students that represented the range of computational fluency pretest scores within one second-grade classroom and balanced gender in the sample. Two students (one male, one female) with pretest scores in the upper quartile, two students (one male, one female) with pretest scores in the lower quartile, and two students (one male, one female) with pretest scores near the median were selected for interviews. Additionally, to be included as an in-depth case, only students who returned the informed consent with the student and parent/guardian consenting to the videotaped interviews were selected.

### *Data Sources*

Data were collected using paper-pencil assessments and semi-structured interviews at four time points over nine weeks, which are described as pretest (Week 0), benchmark 1 (Week 3), benchmark 2 (Week 6), and posttest (Week 9). The paper-pencil assessments consisted of two main assessment instruments: 1) The Assessment of Math Fact Fluency, and 2) Word Problem Situations. The pretest and posttest were identical. The benchmark tests differed in terms of varying the order of the computation problems on the Assessment of Math Fact Fluency problems and varying the numbers for the same problem-types on the Word Problem situations section (e.g., instead of 43 and 12 used in the separate-result-unknown problem type, 37 and 10 were used for benchmark 1 and 43 and 10 were used in benchmark 2).

The Assessment of Math Fact Fluency (Fuchs et al., 2003) section of the assessments was a battery of addition and subtraction problems that measure computational efficiency and accuracy. Students had one minute on each page of four fluency measures, with each page containing 25 addition or subtraction problems (sums up to 18 and minuends up to 18). The coefficient alpha for calculation fluency in third grade, on a tested sample, was equal to or greater than .89 for each subtest (Fuchs et al., 2003; Locuniak & Jordan, 2008). There were 100 computation problems total with 50 addition problems and 50 subtraction problems. The total score was calculated as the percentage correct out of 100 total problems.

The Word Problem Situations section of the assessments included four different Cognitively Guided Instruction problem types (Carpenter et al., 2015). Multiplication, part-part-whole, separate-result-unknown, and join-change-unknown problem types were used for all the assessments in order to measure students' computation of one- and two-digit numbers within a word problem context. Unlike the Assessment of Math Fact Fluency, this portion of the assessment was not timed, and students were encouraged to solve the problems in ways that made sense to them and to show their thinking. There were four word problems total. The total score was calculated as the percentage correct out of 4 total problems.

The researchers interviewed the participants in four semi-structured interviews about their strategies for solving the fact fluency and word problem assessments after taking the paper-pencil assessment. During the interviews, students were asked to explain how they solved specific computation problems and word problems in order to elicit students' strategies and reasoning for solving problems (Cai, 2000). Specifically, students were asked questions such as *How did you solve the problem?*, *Tell me what you did here to figure it out*, and *Can you tell me how you counted it?*

### *Data Analysis*

To explore the variations in students' achievement and strategies, data were analyzed using descriptive statistics (quantitative), graphical representations (quantitative), open and axial coding (qualitative), and frequency matrices (quantitative and qualitative).

*Quantitative analysis.* The paper-pencil assessments were the primary data sources for the quantitative analyses. Descriptive statistics included measures of central tendency and indicators of dispersion to provide an overview of students' performance on the measures. We used boxplots (across subjects) and line graphs (within subjects) to examine the shape of the data, visualize unusual aspects of the data, and graphically interpret the dispersion of the data. These analyses summarized and aided us in making sense of the data (Cohen, 2008).

*Qualitative analysis.* The semi-structured interviews were the primary data sources for the qualitative analyses. Preliminary analysis entailed an examination of each student's video-recorded interview and annotations of initial descriptions and interpretations of the student's verbal explanations, actions, and behaviors. We then used Baroody's (2006) phases of basic fact fluency (i.e., 1 - Counting Strategies, 2 - Reasoning Strategies, 3 - Mastery) to code students' strategies on the Assessment of Math Fact Fluency and the Cognitively Guided Instruction strategy types (i.e., direct modeling, counting, and retrieval; Carpenter et al., 2015) on the Word Problems. Within these phases and strategy types, we then transcribed selected portions of the interviews and used axial coding to further sort and define themes within the phases of strategy development (Miles & Huberman, 1994). Axial codes for computational strategies were used, such as "make a ten," "compensation," and "doubles-plus-one." Axial codes for use of number sense were also used, in alignment with the Active Construction View, such as "decomposed numbers," "used place value language," and "indicated sense of magnitude" (Baroody et al., 2009). The codes were used to discern themes, patterns, and processes and make comparisons (Coffey & Atkinson, 1996). The process involved multiple iterations of analyzing the interview data.

*Mixed analysis.* Mixed quantitative and qualitative analyses occurred after each of the strands were analyzed separately. A visual analysis of the line graphs for each measure from each student provided a trajectory analysis of the shape of data over time and an analysis comparing similar and different shapes across students (i.e. increasing, peak and decrease, and flat) and student categories (i.e. students with pretest scores in the upper or lower quartiles or near the median). Rather than use a linear aggregate, we plotted line graphs for each measure (i.e. the Assessment of Math Fact Fluency and the Word Problem Situations) in order to preserve their discrete meaning and better understand the nuances of the data.

Matrices were used as a visualization tool for analyzing trends and pinpointing divergence and convergence in the data. The matrices were organized by student categories (upper quartile, near median, lower quartile) and slope patterns (increasing, peak and decrease, and flat). We qualitized our quantitative data by organizing the slope and test score patterns by theme, analyzing the trends holistically, and then comparing and contrasting the matrices. Within the dialectics perspective, we were especially interested in points where the data diverged and used both the quantitative and qualitative data to understand these variations. This allowed us to examine subtle shifts in students' computational fluency, including accuracy and proficiency on the paper-pencil measure and flexibility and appropriateness of strategies in the interviews.

## Results and Discussion

The purpose of this study was to explore the variations in students' achievement scores and strategies on computational fluency assessments and interviews. The results are reported under four main headings: 1) Variation Within Students' Paper-Pencil Assessments, 2) Variation Within Students' Interviews, 3) Variation in Students' Trajectories: The Convergence and Divergence of Data, and 4) Variation in Students' Trajectories: Considering Students' Starting Points. The first section focuses on the line graphs of achievement scores and matrices (mixed analyses) of variation in students' scores. The second section presents and describes the qualitative analysis of variation in students' strategies for solving computation problems and the ways in which the interview results converged, diverged, and or/or augmented the assessment score results (mixed analyses). The final two sections present the results and interpretation of the variation in students' trajectories in terms of overall themes in the convergence/divergence of data and in terms of students' increases/decreases in scores and strategies in terms of their starting point (i.e., pretest scores).

### *Variation Within Students' Paper-Pencil Assessments*

*Assessment of Math Fact Fluency.* Figure 2 presents the line graphs of each student's paper-pencil assessment scores for the fact fluency measure at each measurement point. Arthur and Beth scored in the upper quartile on the pretest; Cathy and Dillon scored near the median on the pretest; and Ed and Frankie scored in the lower quartile on the pretest. While the line graphs highlight these differences in pretest achievement (i.e., student categories of upper quartile, near the median, and lower quartile), they also show that each student's trajectory across the four measures was different. All students, except Ed, showed an increase in their fact fluency score at two or more of the measurement points. However, only Arthur, Beth, and Cathy showed an increase from pretest to posttest (and Arthur had a peak score prior to the posttest).

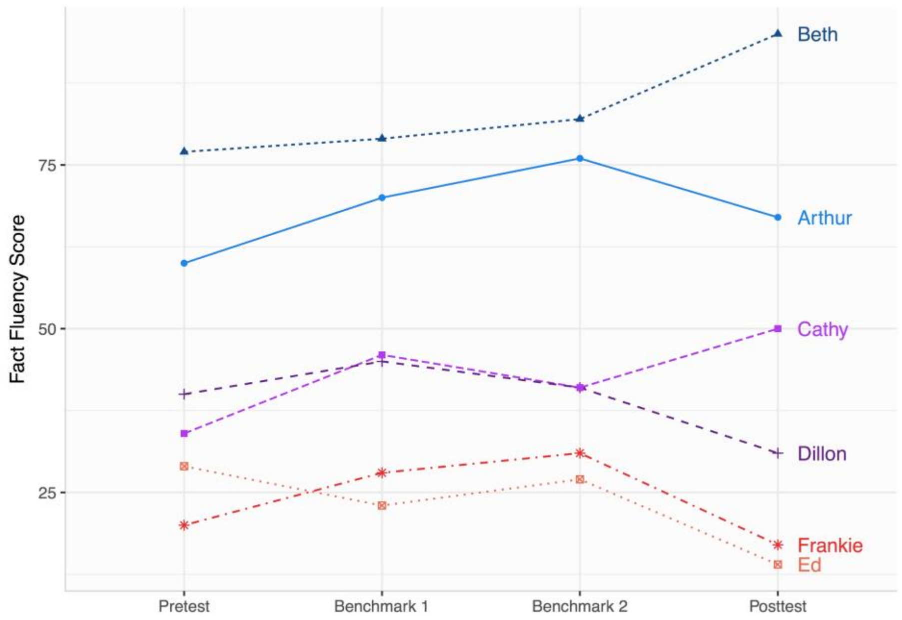


*Figure 2.* Variations in students' Assessment of Math Fact Fluency scores.

Table 1 displays a matrix of the shape patterns in the line graphs for the fact fluency measure. Beth is the only student that had a consistent increase in scores across measurement time points. Arthur had a consistent increase in scores until posttest, which decreased by 9 points (categorized as "Peak and Decreasing"). Cathy and Dillon, both students who had a pretest score near the median, exhibited a relatively flat trajectory. These were categorized as flat because the scores remained within a range of 10 points across measurement points (i.e., for Cathy, within 41 to 50 from benchmark 1 to posttest; for Dillon, within 40 to 45 from pretest to benchmark 2 with a decrease to 31 on the posttest). Ed and Frankie, the two students with low pretest scores, had a peak in scores, followed by a decrease (i.e., for Ed, a decrease of 13 points at posttest; for Frankie, a decrease of 14 points at posttest).

Table 1
*Shape Patterns for Assessment of Math Fact Fluency Scores*

| Student Category | Increasing | Peak and Decreasing | Flat | Total |
|---|---|---|---|---|
| Upper quartile | 1 | 1 | | 2 |
| Near median | | | 2 | 2 |
| Lower quartile | | 2 | | 2 |
| | | | | |
| Total | 1 | 3 | 2 | 6 |

*Word Problem Situations.* Figure 3 presents the line graphs of each student's assessment scores for the word problem situations measure at each measurement point. Dillon exhibited a ceiling effect, as he scored 100% at each measurement point. Beth, similar to her trajectory for fact fluency, again showed consistent increase in scores at each measurement point. While Arthur and Cathy did not show consistent increases on the fact fluency measure, they did show a consistent increase in scores on the word problem situations measure. Ed and Frankie, similar to the pattern of their fact fluency scores, again exhibited a peak in their test scores for the word problem situations followed by a decline in their score.

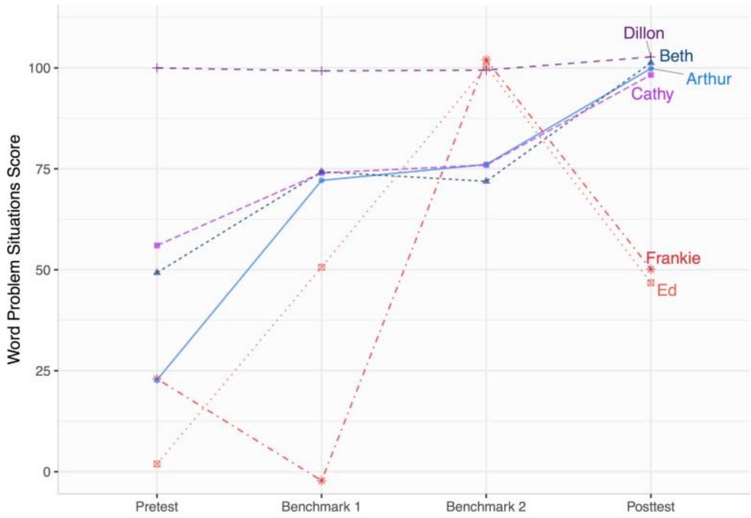


*Figure 3.* Variations in students' Word Problem Situations assessment scores.

Table 2 displays a matrix of the shape patterns in the line graphs for the word problem situations measure. Again, students were categorized by their fact fluency pretest score. Arthur and Beth both had increases over the four measurement points. In terms of the shape pattern, Cathy and Dillon's trajectories were different on this measure. While both students had relatively flat trajectories on the fact fluency measure, Cathy's shape pattern was increasing on the word problem situations measure. While Dillon's was again categorized as flat, this time it was due to high scores across the four measures. Ed and Frankie, similar to the shapes of their fact fluency trajectories, again had a peak and a decrease in scores. The shape pattern for their word problem situations measure mirrored the inconsistent scores in their shape pattern for the fact fluency measure across the four measurement points.

Table 2
*Shape Patterns for Word Problem Situations Scores*

| Student Category | Increasing | Peak and Decreasing | Flat | Total |
|---|---|---|---|---|
| Upper quartile | 2 | | | 2 |
| Near median | 1 | | 1 | 2 |
| Lower quartile | | 2 | | 2 |
| Total | 2 | 2 | 1 | 6 |

### *Variation Within Students' Interviews*

Similar to the trajectory analysis of assessment scores, the qualitative analysis of students' strategies for solving computation problems revealed that each student had a different trajectory across the nine weeks of the study. The mixed analysis provided insights about broad trends in these data and uncovered divergent results in the data. This section is organized by measure: first, with the results of the fact fluency measure; second, with the results of the word problem situations measure.

*Assessment of Math Fact Fluency.* Beth, who had consistently increasing fact fluency scores throughout the study, displayed a variety of strategies on the pretest interview, followed by the consistent use of two main strategies (make a ten and count down) on the benchmarks and posttest. This pattern in her strategies combined with an analysis of her assessment scores shows that, over time, she became more efficient with these Phase 1 – Counting and Phase 2 - Reasoning strategies and increased her speed and accuracy. The quantitative and qualitative results converged and indicated steady growth in accuracy and efficiency of math fact fluency and consistent use of two strategies (Phase 1 – Counting and Phase 2 – Reasoning) across measurement points.

Arthur and Dillon did not display as many strategies as Beth on the pretest. However, both Arthur and Dillon showed a change from pretest to benchmark 1, where their variety of Phase 2 – Reasoning strategies increased and this variety of strategies was maintained throughout the benchmark 2 and posttest interviews. In looking at their line graphs of fact fluency assessment scores, both students had an increase in scores at benchmark 1, the point in which they accessed a variety of strategies to solve the problems. For example, instead of only using the "make a ten" strategy (Phase 2 – Reasoning) as he did on the pretest, on benchmark 1 Arthur used a variety of reasoning strategies such as "make a ten," "number relationships," and "constant difference" strategies. While Arthur improved in explaining these strategies during the posttest interview, his fact fluency test score decreased by nine points from benchmark 2 to posttest. This was a divergence in the quantitative and qualitative data.

A similar pattern occurred in a comparison of Dillon's test score and interview data. During the benchmark 1 interview, Dillon began using Phase 3 – Mastery "retrieved facts" and Phase 2 – Reasoning strategies "double-minus-one," and "compensation." His fact fluency assessment score had also increased. However, Dillon's fact fluency test scores decreased after benchmark 1, despite his improved fluency with these strategies during each subsequent interview. Similar to Arthur, Dillon's quantitative and qualitative data showed convergence at benchmark 1, but then seemed to diverge at benchmark 2 and posttest.

Arthur and Dillon's results showed a convergence of test score data and interview data at benchmark 1, when their test scores improved along with the presence of more strategies for solving the fact fluency problems. Their quantitative and qualitative data diverged, however, after benchmark 1 (at posttest for Arthur and at benchmark 2 for Dillon). Cathy's results also diverged, but did so differently. Cathy's test score and interview results diverged at benchmark 2, a point at which her test score decreased with the presence of more strategies. Initially, Cathy relied on Phase 1 – Counting strategies using her fingers, and a few Phase 3 – Mastery memorized facts throughout the pretest and benchmark 1 interviews. During benchmark 1, she exhibited two instances in which she looked for combinations of ten to solve the problems more easily (Phase 2 – Reasoning). At benchmark 2, her strategies changed a great deal as she began to more frequently use Phase 2 – Reasoning strategies such as doubles facts as an anchor (e.g., doubles-plus-one strategy) and number relationships. This change in Cathy's variety of reasoning strategies corresponded with a drop in her assessment score for computational fluency. However, her assessment score went up on the posttest. The interview data showed that Cathy maintained the use of a variety of Phase 2 - Reasoning strategies on the posttest, though she continued to rely on Phase 1 - Counting for problems she perceived to be difficult.

Ed's trajectory mirrored Cathy's trajectory in some ways, though his was a more gradual shift in using strategies based in number sense (i.e., the Active Construction View; Baroody et al., 2009). On benchmark 1, Ed showed a decrease in his fact fluency score. In the interview, Ed attempted new Phase 2 – Reasoning strategies, such as doubles-plus-one and compensation, both of which provided insight into Ed's engagement of number sense. For example, as he used doubles-plus-one and compensation strategies, he decomposed numbers, used anchor numbers, and described relationships among numbers. At benchmark 2, Ed's assessment score increased, and he added more reasoning strategies to his repertoire. Ed maintained his variety of Phase 2 – Reasoning strategy use on the posttest, although his assessment score dropped down to 14% (from 27% at benchmark 2).

Finally, Frankie's pretest and benchmark interviews showed her reluctance to use her number sense to solve fact fluency problems. In these interviews, she relied on Phase 1 - Counting strategies, with an occasional use of a Phase 3 – Mastery memorized fact. Her assessment scores increased at the pretest and benchmark measurement points. During the posttest interview, Frankie exhibited two instances of the Phase 2 – Reasoning strategies of "make a ten" and four instances of the "number relationships" reasoning strategies. Her verbal explanation of these strategies showed growth in her number sense knowledge as she was able to decompose numbers and use relationships among numbers to solve the problems more fluently. These are key characteristics of the flexibility and appropriate use of solution strategies. This improvement in reasoning strategy use corresponded with a decline in her assessment score, another divergent finding, and very similar to the divergence in data exhibited in Ed's trajectories over time.

*Word Problem Situations.* Both Arthur and Beth showed consistent increases in their word problem situations assessment scores as well as consistency in their strategies across measurement points. In both cases, the test score data converged with the interview data for the word problem situations. Arthur's pretest score on this measure was low (25%), but he was able to quickly describe and solve each problem during the interview. The strategies he explained in the interview remained consistent throughout the subsequent interviews. He often used pictorial representations of base ten blocks to highlight his thinking, but in the interviews, he explained that he did not need the pictures to solve the problem.

Beth's pretest score on this measure was higher than Arthur's (50%), but unlike Arthur, she struggled to explain her strategies and used simple counting strategies rather than strategies based in number sense. At benchmark 1, Beth began using equations and algorithms to show her thinking. As she described her strategies, Beth's engagement of number sense was clear as she was able to verbally describe numbers' decomposition into tens and ones and the use of relationships among place value positions to solve the problems. While her descriptions became more algorithmic in nature at benchmark 2 and posttest, evidence of her number sense was present as she referred to the values of digits (i.e., "30" instead of "3" in the number 34) as she described her procedures for solving the problems.

Similar to Arthur and Beth, Cathy also showed a consistent increase in her word problem situations assessment scores. Unlike Arthur and Beth, Cathy's strategies were not consistent, and her strategies often did not exhibit engagement of number sense. While her accuracy on the assessments scores improved over the nine weeks, her strategies for solving the problems did not improve, an interesting divergence in the data. Cathy relied on counting in most interviews. Interestingly in benchmark 1, she used place value language on two of the problems (i.e., described tens and ones); however, she regressed back to counting during the benchmark 2 and posttest interviews.

While there was a peak and decrease in both Ed and Frankie's assessment scores for the word problem situations, both students showed the most growth among the six cases in terms of using strategies based in number sense. Ed's scores were inconsistent, but during the interviews, he could correctly solve all of the problems. During the pretest interview, Ed was allowed to use tools (such as the cubes or hundreds chart) to solve the problems for which he originally had no solution, and correctly solved each problem. During the benchmark 1 interview, Ed chose not to use tools. Instead he explained the pictorial representations of the problems (e.g., bags of books; tens and ones blocks) he drew during the assessment and was able to find a solution to the problems during the interview. While his strategies for solving the problems were coded as a lower-level strategy (i.e., direct modeling), his explanations revealed number sense understandings such as the decomposition of two-digit numbers into tens and ones and combining groups of numbers through skip counting number relationships. In the posttest interview, Ed began connecting symbolic notations to his drawings and was able to explain the composition of numbers and number relationships. His assessment score dropped on the posttest because he was missing a solution to one problem, but during the interview he easily explained his strategy and solution. In these kinds of cases, the results of the quantitative and qualitative data seem divergent; however, a careful analysis of the qualitative data reveals what Ed actually knows and augments the assessment score data. In most of these cases, Ed needed to verbally describe his mathematical ideas and had difficulty completing the tasks independently.

Frankie relied on drawing pictorial representations to solve the word problem situations throughout each of the four interviews. However, there were important changes in her strategies over the course of the interviews. In the pretest interview, Frankie did not attempt two of the problems and maintained her incorrect answer on another problem. Frankie attempted all four problems on benchmark 1, but scored 0%. Nonetheless, during the benchmark 1 interview, she explained her drawings and provided a correct solution on 75% of the problems. Her verbal explanations included place value language and the decomposing of numbers. At benchmark 2, Frankie still relied on drawing pictorial representations both to help her understand the problem and solve the problem, but she also began connecting these pictorial representations to symbolic notations. Her explanation of the equations tied to the pictorial representations demonstrated a linking between her sense of quantities and symbolic number sense. During the posttest interview, Frankie continued to demonstrate this change in her thinking on two of the problems, yet she also struggled with solving the last two problems in the assessment and reverted back to counting by ones to solve them.

Dillon's trajectory on the word problem situations assessment was distinctly different from the trajectories of the other five students. Dillon scored 100% on all four assessments. Similarly, the codes for strategies and number sense were also consistent throughout the interviews. The one change that took place during his interviews was on the posttest interview in which he used an algorithm to solve two problems. But even with this change in strategy, his explanation of the algorithm procedure maintained his number sense codes for using place value language and decomposing numbers. Dillon's consistent results in his test scores and interviews showed a strong convergence of the data. At the same time, his use of and explanation about an algorithm in the posttest interview augmented the test score data. This shift in Dillon's strategy use would have been hidden in the assessment score were it not for the interview data.

*Variation in Students' Trajectories: The Convergence and Divergence of Data*

The integration of qualitative and quantitative data helped us better understand what occurred as students' learning moved within Baroody's (2006) three phases of computational fluency. For example, the qualitative and quantitative data converged in Beth's case. The data from Beth's test scores showed steady growth in accuracy and efficiency across measurement points, and her interview data about her strategies indicated consistent use of two main strategies from benchmark 1 to posttest. From the pretest to benchmark 1, Beth's strategies were moving beyond Phase 2 – Reasoning Strategies. The evidence from benchmark 1 to posttest showed that Beth's strategies were situated in Phase 3 - Mastery. Overall, her growth during the nine-week study was in speed and accuracy. The reason for this could be that Beth began the study with an already robust number sense (Baroody & Rosu, 2006).

The qualitative and quantitative data converged in the cases of Arthur and Dillon at the benchmark 1 measurement point. Strategy use and test scores increased at benchmark 1. However, these data diverged for Dillon at benchmark 2 and posttest. As Dillon improved in explaining a variety of reasoning strategies, his test scores decreased. Similarly, this occurred with Arthur at the posttest. This could mean that as Arthur and Dillon developed their reasoning strategies and used them more appropriately and flexibly, it might have taken them more time and/or effected their accuracy. While their test scores could look concerning and show not only a lack of growth, but a decrease in achievement, their interview data highlighted an important phase in their computational fluency development. They needed time to work through developing reasoning strategies based in number sense (Jordan, Kaplan, Ramineni, & Locuniak, 2008; Jordan, Glutting, & Ramineni, 2010).

This divergence in the qualitative and quantitative data was also evidenced in Cathy, Ed, and Frankie's cases. For instance, at the benchmark 2 measurement point, Cathy's interview data indicated an increase in the variety of strategies and an improvement in terms of reasoning strategies based in number sense. This improvement corresponded with a decline in her test score, much like what occurred in Arthur and Dillon's cases. Her trajectory to the posttest allowed us to see an increase in her test score but lack of growth in terms of strategies. This could mean that Cathy was maintaining her variety of strategies and improving her use of them, which translated into improved accuracy and efficiency. In fact, her interview augmented the test score data showing that she maintained the use of strategies on the posttest (which had been new at benchmark 2) and she relied on counting for problems she perceived as difficult. This places Cathy's strategy development in Baroody's (2006) Phases 1 - Counting and 2 - Reasoning Strategies. At the posttest point in time, Cathy was solidifying new reasoning strategies yet not introducing new strategies for difficult problems. Again, the test score data alerted us to inconsistencies in the scores across measurement points; the interview data provided insights into Cathy's subtle shifts in learning. Counting strategies could be the starting point for helping students like Cathy develop strategies based in number sense. However, reliance on counting strategies can hinder progress over time. In a longitudinal study of kindergarten to second-grade students, Jordan et al. (2008) examined the change in students' frequency of finger counting on number combinations in relation to change in their accuracy. Jordan et al. (2008) found that finger counting was positively correlated with accuracy in kindergarten, but by the end of second grade there was a significant negative correlation between finger counting and accuracy.

Frankie's case was similar to Cathy's in that when Frankie used counting strategies that were familiar to her (i.e. did not try new reasoning strategies) her test scores improved (benchmark 1). But, as she tried new reasoning strategies on the posttest, moving her strategies into Phase 2, her test scores decreased. This could mean that learning new strategies slowed down her computation and/or affected her accuracy. This divergence in the data provided a more complete picture of her improvements in accuracy and efficiency, as well as improvements in strategy development. Based on the Jordan et al. (2008) study, one could view this shift to Phase 2 strategies, despite the decrease in test scores, as a positive development for Frankie's computational fluency in the long run.

Ed's divergent data painted a similar picture, but his interview data augmented the test score data. At benchmark 1, Ed's interview showed an improvement in strategy usage, but his test scores decreased. His test scores improved at benchmark 2, but then decreased again at posttest. Simultaneously, he acquired some new reasoning strategies at benchmark 2 (the qualitative and quantitative data converged) and while he maintained flexible and appropriate use of strategies at the posttest, his test scores showed a decline in accuracy and efficiency (divergence). This led us to infer that Ed was working within Phase 2 - Reasoning Strategies; the convergence and divergence between the data at different measurement points evidenced this inconsistent interplay among accuracy, efficiency, flexibility, and appropriately using strategies. And because Ed showed the most growth in reasoning strategies over the nine weeks, it is possible we witnessed a critical time in the development of his computational fluency. He seemed to be developing understanding within Phase 2 of basic fact fluency.

### *Variation in Students' Trajectories: Considering Students' Starting Points*

Overall, there was variation in students' assessment scores, strategy use, and engagement of number sense, no matter their starting point at the pretest. However, visualizing the data within student categories based on pretest score (low, median, and high) through line graphs and matrices also provided some insight into students' computational development. On the fact fluency measure, Arthur and Beth, both students with high pretest scores, did not have as much variation in their strategies as the other four students. Dillon and Cathy, students with pretest scores near the median, showed slightly more variation in their strategies than Arthur and Beth during their interviews. Dillon seemed to solidify his strategy use during the time between benchmarks 1 and 2. By the posttest interview Dillon's strategies were less varied, and he consistently relied on a set of efficient strategies based in number sense. At the posttest, Dillon's strategies and approaches to the problems looked more like Beth's strategies. This means that Dillon transitioned from Baroody's Phase 2 to Phase 3 during the nine weeks of the study. However, in only looking at the line graphs, his test score indicated that he was performing far below Beth. The combination of data sources highlighted the need for Dillon to improve his accuracy and efficiency, while the interview data provided insight into his movement among fact fluency phases and the progress he made with flexibility and appropriate use of strategies.

Ed and Frankie, students with low pretest scores, showed the most growth in flexibility and appropriate use of strategies when compared to the other four students, although their assessment scores did not reflect this trend. For example, Ed revealed correct mathematical reasoning based in number sense throughout the word problem situation interviews, but scored low on this measure in three out of the four assessments. One possible reason for this divergence in the quantitative and qualitative data analysis could be due to a heavier demand on working memory (Geary et al., 2004) as he used new knowledge to solve new problems. His use of number sense knowledge within his strategies improved over the nine weeks. From the perspective of the Active Construction View, Ed's computational fluency remained in Phase 2. But, because he did not skip Phase 2, it is possible that Ed's improvement in strategy use and flexibility could lead him to more accurate and efficient procedures on assessment scores as this interconnected knowledge becomes more automatic (Baroody et al., 2009; Baroody et al., 2007; Jordan et al., 2010).

## Limitations

We used a mixed-methods design in order to draw from the strengths of both quantitative and qualitative data analyses. However, the mixed methods approach and the design for this study had limitations. We were limited in our quantitative analysis (i.e., relied on descriptive statistics, not inferential statistics) due to the small sample size of six students. A limited sample, however, allowed us to analyze students' thinking processes in more depth.

The small sample also led us to use a purposive sample rather than a random sample for the six cases. While a probability sample has benefits, it would have overlooked aspects of pretest knowledge that we wanted to investigate. We used students' pretest scores to select a range of students from students with low pretest scores and scores near the median to students with high pretest scores. An additional limitation was that all students were from the same second-grade classroom, which limited the conclusions in terms of generalizability to other schools or contexts.

## Conclusion

The results showed that these six students' variations in computational fluency were evident in their starting points (i.e., pretest scores) and trajectories over the nine weeks. Students with lower pretest scores showed changes in strategy use (often from Phase 1 to 2), though some still relied on Phase 1 strategies for problems they perceived as difficult. Students with median-range pretest scores showed more variation in Phase 2 – Reasoning strategies. Students who scored higher than peers on the pretest already had an established range of "go to" strategies that they relied on for efficiency and accuracy, often moving more directly into the Phase 3 – Mastery of retrieval strategies.

The results also showed variations in computational fluency as students' interview data converged with, diverged from, or augmented their assessment data, which provided further insight into each students' computational fluency over the nine weeks. For example, Beth's strategies on both the fact fluency and word problem measures did not vary, yet her assessment scores improved over the nine weeks. Frankie's assessment scores dropped at the posttest measurement point, but her use of strategies based in number sense improved throughout the interviews. In both these examples, the students' assessment score trajectories looked different from the trajectories of their strategy growth and development exhibited in the interviews. Overall, in five of the six cases (i.e., Arthur, Cathy, Dillion, Frankie, and Ed), improvement in reasoning strategies corresponded with a decline in the achievement scores. Our conclusions suggest that the quantitative test score data or the qualitative interview data alone would have been limited in scope, especially with regard to the current notion of computational fluency as a construct that involves accuracy, efficiency, flexibility, and appropriate use of strategies.

### *Implications for Mathematics Teachers*

The implication for mathematics teachers of seven- and eight-year-old students is to understand these phases of computational fluency as it relates to current (more expanded) definitions of the construct. It is important for teachers to understand that typical fact fluency test scores by themselves will not reflect students' growth in strategy development. The results of this study suggest that students' computational fluency achievement scores often decline as they move through the Phase 2 – Reasoning stage of learning. With a variety of computational experiences, students in this age group gradually acquire more fluency with reasoning strategies and begin their movement into the Phase 3 – Retrieval, which often results in a new increase in computational fluency achievement scores, much like Beth's trajectory.

## References

Bamberger, M., Rugh, J., & Mabry, L. (2012). *Real world evaluation: Working under budget, time, data, and political constraints* (2nd ed.). Thousand Oaks, CA: Sage Publications, Inc.

Baroody, A. J. (2006). Why children have difficulties mastering the basic fact combinations and how to help them. *Teaching Children Mathematics*, *13*(1), 22-31.

Baroody, A. J., Bajwa, N. P., & Eiland, M. (2009). Why can't Johnny remember the basic facts? *Developmental Disabilities Research Reviews, 15*, 69-79.

Baroody, A. J., Feil, Y., & Johnson, A. R. (2007). An alternative reconceptualization of procedural and conceptual knowledge. *Journal for Research in Mathematics Education, 38*(2), 115-131. Retrieved from: http://www.jstor.org/stable/30034952

Cai, J. (2000). Mathematical thinking involved in U.S. and Chinese students' solving of process-constrained and process-open problems. *Mathematical Thinking and Learning, 2*(4), 309-340.

Carpenter, T. P., Fennema, E., Franke, M. L., Levi, L., & Empson, S. B. (2015). *Children's mathematics: Cognitively guided instruction, 2nd ed*. Portsmouth, NH: Heinemann.

Coffey, A. & Atkinson, P. (1996). *Making sense of qualitative data: Complementary research strategies.* Thousand Oaks, CA: Sage.

Cohen, B. H. (2008). *Explaining psychological statistics* (3rd ed.). Hoboken, NJ: John Wiley & Sons, Inc.

Common Core State Standards Initiative (CCSSI). (2010). *Common core state standards for mathematics*. Washington, DC: National Governors Association Center for Best Practices and the Council of Chief State School Officers. Retrieved from http://corestandards.org

Creswell, J. W., & Plano Clark, V. L. (2011). *Designing and conducting mixed methods research* (2nd ed.). Thousand Oaks, CA: Sage Publications, Inc.

Erlwanger, S. H. (1973). Benny's conception of rules and answers in IPI mathematics. In T. Carpenter, J. Dossey, & J. Koehler (Eds.), *Classics in mathematics education research* (pp. 48- 59). Reston, VA: The National Council of Teachers of Mathematics.

Fuchs, L. S., Hamlett, C. L., & Powell, S. R. (2003). *Grade 3 math battery.* (Available from L. S. Fuchs, 328 Peabody, Vanderbilt University, Nashville, TN 37203)

Geary, D. C. (2011). Cognitive predictors of achievement growth in mathematics: A 5-year longitudinal study. *Developmental Psychology, 47*(6), 1539–1552.

Geary, D. C. (2013). Early foundations for mathematics learning and their relations to learning disabilities. *Current Directions in Psychological Science, 22*(1), 23-27.

Geary, D. C., Hoard, M. K., Byrd-Craven, J., & DeSoto, M. C. (2004). Strategy choices in simple and complex addition: Contributions of working memory and counting knowledge for children with mathematical disability. *Journal of Experimental Child Psychology*, *88,* 121-151. doi: 10.1016/j.jecp.2004.03.002

Göbel, S. M., Watson, S. E., Lervag, A., & Hulme, C. (2014). Children's arithmetic development: It is number knowledge, not the approximate number sense, that counts. *Psychological Science*, 1-10. doi: 10.1177/0956797613516471

Johnson, B., & Gray, R. (2010). A history of philosophical and theoretical issues for mixed methods research. In A. Tashakkori & C. Teddlie (Eds.), *Sage handbook of mixed methods in social and behavioral research* (pp. 69-94). Thousand Oaks, CA: Sage Publications, Inc.

Jordan, N. C., Glutting, J., & Ramineni, C. (2010). The importance of number sense to mathematics achievement in first and third grades. *Learning and Individual Differences, 20*, 82-88.

Jordan, N. C., Hanich, L. B., & Kaplan, D. (2003). Arithmetic fact mastery in young children: A longitudinal investigation. *Journal of Experimental Child Psychology, 85*, 103-119. doi:10.1016/S0022-0965(03)00032-8

Jordan, N. C., Kaplan, D., Ramineni, C., & Locuniak, M. N. (2009). Early math matters: Kindergarten number competence and later mathematics outcomes. *Developmental Psychology*, *45*(3), 850–867.

Jordan, N. C., Kaplan, D., Ramineni, C., & Locuniak, M. N. (2008). Development of number combination skill in the early school years: when do fingers help? *Developmental Science*, *11*(5), 662–668. doi:10.1111/j.1467-7687.2008.00715.x

Lesh, R., Lovitts, B., & Kelly, A. E. (2000). Purposes and assumptions of this book. In A. Kelly & R. Lesh (Eds.), *Handbook of research design in mathematics and science education*. Mahwah, NJ: Lawrence Erlbaum Associates, Inc.

Miles, M. B., & Huberman, A. M. (1994). *An expanded sourcebook: Qualitative data analysis* (2nd ed.). Thousand Oaks, CA: Sage Publications, Inc.

National Council of Teachers of Mathematics. (2000). *Principles and standards for school mathematics.* Reston, VA: National Council of Teachers of Mathematics.

Niglas, K. (2010). The multidimensional model of research methodology: An integrated set of continua. In A. Tashakkori & C. Teddlie (Eds.), *Sage handbook of mixed methods in social and behavioral research* (pp. 215-236). Thousand Oaks, CA: Sage Publications, Inc.

Plano Clark, V. L., Anderson, N., Wertz, J. A., Zhou, Y., Schumacher, K., & Miaskowski, C. (2015). Conceptualizing longitudinal mixed methods designs: A methodological review of health sciences research. *Journal of Mixed Methods Research, 9*(4), 297-319. doi: 10.1177/1558689814543563

Schoenfeld, A. H. (2004). The math wars. *Educational Policy, 18*(1), 253–286. doi:10.1177/0895904803260042.

Shannon-Baker, P. (2016). Making paradigms meaningful in mixed methods research. *Journal of Mixed Methods Research, 10*(4), 319-334. doi:10.1177/1558689815575861

Siegler, R. S. (2000). The rebirth of children's learning. *Child Development, 71*(1), 26-35. doi: 0009-3920/2000/7101-0004

Siegler, R. S., & Shrager, J. (1984). Strategy choices in addition and subtraction: How do children know what to do? In C. Sophian (Ed.), *The origins of cognitive skills* (pp. 229-293). Hillsdale, NJ: Erlbaum.

Taylor, K. L., & Dionne, J. P. (2000). Accessing problem solving strategy knowledge: The complementary use of concurrent verbal protocols and retrospective debriefing. *Journal of Educational Psychology, 92*(3), 413-425.

Vasilyeva, M., Laski, E. V., & Shen, C. (2015). Computational fluency and strategy choice predict individual and cross-national differences in complex arithmetic. *Developmental Psychology, 51*(10), 1489-1500. doi: http://dx.doi.org/10.1037/dev0000045