Using Think-Aloud Protocols to Uncover Misconceptions and Improve Developmental Math Instruction: An Exploratory Study

Charles Secolsky
Alternative Assessment Strategies, csecolsky@gmail.com

Thomas P. Judd
United States Military Academy, thomas.judd@usma.edu

Eric Magaram
SUNY Rockland Community College, emagaram@sunyrockland.edu

Stephen H. Levy
Saint Peter's University, drslevy@gmail.com

Bruce Kossar
Independent Consultant, bruce@brubumski.com

See next page for additional authors

Using Think-Aloud Protocols to Uncover Misconceptions and Improve Developmental Math Instruction: An Exploratory Study

Abstract

Deficiencies in education continue to escalate around the world. The focus on outcomes assessment has narrowed instructional research and curriculum evaluation to standardized testing in certain subject areas. A prototype for a quantitative literacy assessment instrument was developed with the goal of diagnosing student misconceptions of basic mathematics content and changing instructional practices to undo the misconceptions by applying cognitive psychological theory. Two hundred thirty-eight basic math high school students and 209 remedial community college students in New Jersey and New York were administered the instrument, which had been based on coded data from think-aloud protocols. The instrument asked students to answer 20 basic mathematics items and, in addition, to evaluate four possible solution strategies. For each item, frequencies of selected solution strategies and the association between strategy selection and performance on the 20-question math test are presented as a means for improving instruction. Follow-up research is proposed for determining whether undoing the student misconceptions first before teaching material on a new unit of instruction may yield more positive student outcomes.

Keywords

quantitative literacy assessment, basic mathematics, instructional development

Creative Commons License

Cover Page Footnote

Charles Secolsky specializes in educational measurement, applied statistics, and program evaluation. He is president of Alternative Assessment Strategies Inc.
Thomas P. Judd is Assistant Dean of Academic Assessment and Associate Professor at United States Military Academy.
Eric Magaram is Assistant Professor and Chair of Mathematics at Rockland Community College
Stephen H. Levy is Professor of Philosophy at Saint Peter's University and Adjunct Professor of Bioethics at Holy Name Medical Center.
Bruce Kossar specializes in latent class models and market research and is an independent consultant in Kinnelon, New Jersey.
George C. Reese is Director of Mathematics, Science, Technology in Education at the University of Illinois at Urbana-Champaign.

Authors

Charles Secolsky, Thomas P. Judd, Eric Magaram, Stephen H. Levy, Bruce Kossar, and George Reese

Introduction

For now and for the foreseeable future, there is a need to improve instruction in many parts of the world. Reports of deficiencies in student achievement abound even as demands for accountability continue to grow. For example, consistent with the recommendations of the Spellings Report issued by the U.S. Department of Education (2006), accreditation agencies have increasingly focused on measures of student outcomes. Similarly, in primary and secondary education the No Child Left Behind legislation has generated numerous testing measures all in the name of greater accountability.

As Darling-Hammond (2007) notes, the increased emphasis on producing students with better skills has had unintended consequences. For example, the greater emphasis on testing has ironically led to more, not fewer, children being left behind at least at the lower end of the education spectrum (i.e., high school graduation rates among lower-performing students are now declining). Lower-performing students who do graduate from high school and enroll in community colleges are often placed into developmental courses; the highest frequency of placements occurs in developmental math courses, such as basic math. Unless these students are successful in these basic math courses, they cannot progress to college-level coursework. Unfortunately, very high percentages of students placed in developmental math courses never progress beyond those courses and therefore never realize their goals in higher education. Inability to succeed in basic math courses may present the most significant obstacle to this population's success in college.

The focus on group testing does not improve the quality of education but may instead provide incentives for "dumbing down" the curriculum (Darling-Hammond 2007). Darling-Hammond argues that rather than relying on group-average test scores as the gold standard of accountability, tracking student progress may provide a more valuable and informative measure of accountability that will not only motivate better intellectual work but, as important, will also refocus education on the task of teaching students. This renewed focus on the improvement of teaching would ideally require not only that certified teachers have the necessary knowledge of subject matter but also that these teachers acquire enhanced pedagogical skills that go beyond their essential teaching skills. Generating improvements in pedagogy is one of the primary objectives of research into the teaching-learning process.

This article presents assessment research that, if properly developed and implemented might mitigate learning problems in the classroom and be of value to curriculum developers. First, we present a backdrop and theoretical rationale for the identification of student misconceptions of math content via student interviews known as think-aloud protocols. From the data collected in the think-

aloud protocols, for each of 20 test questions we created a list of four possible solution strategies. A second group of students was asked to complete the math test while simultaneously evaluating the correctness of each potential strategy. This second administration of the test allows us to determine whether discernment of the validity of a particular strategy well-predicts performance on the test. In other words, we are able to identify which misconceptions are most likely inhibiting performance on the exam.

It would seem that using think-aloud protocols to gain an understanding of student misconceptions and their relationship to test performance could productively inform pedagogy. While we have not formally tested this hypothesis by altering curriculum to address misconceptions held by our students, an exploratory analysis produced results consistent with this idea. Two groups of students were given identical math tests. During the exam the treatment group was also asked to evaluate the strategies lists generated in the think-aloud protocol. Even without explicit instruction concerning the misconceptions, the mere invitation for students to evaluate alternative strategies led the treatment group to outperform the control group.

Rationale

Concerns about deficiencies in students' acquisition of critical knowledge and skills during the course of their elementary, secondary, and postsecondary educational experiences in the United States prompted a number of responses by various federal, state, and regional education departments, agencies, and associations (Lacireno-Paquet et al. 2014, Provezis 2010, Schray 2006). Many of these responses can be seen as attempts to provide structure and set standards for learning outcomes, for example, in the form of prescribed curricula and tests and requirements for clearly defined learning outcomes with associated assessment processes.

While structures and standards can be seen as useful for ensuring integrity and equity in educational experiences, they can also result in a narrowing of focus of instructional efforts, both to fulfill requirements and to ensure a positive public presentation. Further, these approaches emphasize the summative end of the teaching-learning-assessment spectrum, perhaps at the expense of important formative information, which is the most timely and thus most useful information for the improvement of the educational process for individual students.

In this regard, optimal strategies would be more formative; those strategies could assist educators in identifying areas of strength or weakness in students' learning of subject matter while the students are in the process of that learning. Corrections applied at key moments would facilitate a student's progress in the acquisition of the targeted knowledge and skill (Askew and Wiliam 1995). Using

these strategies would result in educators' focusing on the educational experience and the quality of that experience. For students, receiving guidance at critical junctures may result in increased understanding, reduced frustration, and increased engagement as well as motivate continued efforts and higher-quality work. These improvements will address more purposefully the underlying concern about educational deficiencies and can help educators and students realize the true objective of greater educational accountability sought by parents, professionals, and accrediting agencies: increased student learning.

This focus on strategies to enhance the education experience would require that teachers acquire new pedagogical skills that go beyond their essential teaching skills. Specifically, the current research explores methods for understanding how students know what they know and uses this knowledge to aid instruction, particularly in learning situations where traditional instructional methods have been deemed to be failing. With a greater understanding of student misconceptions of the skills that are prerequisites to the learning of new material in basic mathematics, teachers may be able to improve student learning before the onset of traditional instruction. If teachers "unteach" or "undo" student misconceptions first, students may become more successful learners with more positive outcomes.

Effective classroom assessment techniques not only provide instructors with evidence of how students are progressing but also uncover the problems encountered on the mastery pathway. While these techniques help the teaching faculty to fine-tune their classroom instruction (Cross and Angelo 1993; Drezek-McConnell and Doolittle 2012), there are deeper levels of understanding about what makes it difficult for students to learn what their instructors are attempting to teach. Bjork's (1994) work on desirable difficulty, whereby a learner ceases to process information because of memory overload, served as the basis for the approach used in this study. His work was the forerunner of cognitive load theory (CLT). Sweller (1988, 1994, 2010) describes intrinsic difficulty as a component of CLT. There is intrinsic cognitive load, and there is extraneous cognitive load. Intrinsic cognitive load represents the number and complexity of the sequential steps needed for correctly completing a task as presented to the student without nonessential material. These steps are in a sense the bare-bones representation of the problem embodied by the task. Extraneous cognitive load, on the other hand, represents those elements of the task that are unrelated to solving the basic problem at hand. Extraneous cognitive load could include how the problem is presented. For example, extraneous cognitive load could be introduced via culturally unfamiliar names of characters in word problems. Extraneous cognitive load can, therefore, mask problem-solving difficulty.

Borrowing from CLT, the research reported herein stems from our position that undoing a misconception can reduce intrinsic cognitive load, especially for

well-defined domains such as with a basic skills assessment. In contrast, Feltovich, Spiro, and Coulson (1993) have argued that for more complicated domains, this reduction of cognitive load would not necessarily occur with the elimination of a misconception because other more advanced problems that require greater cognitive load might arise. In our view, with a developmental population and a basic skills test, many types of misconceptions are likely to occur, each requiring the teacher's attention. The elimination of one or more of these misconceptions may not necessarily reduce the cognitive load in this testing context because of the intrinsic difficulty of the items for these students. Nevertheless, intrinsic item difficulty as part of the assessment instrument then can be used in assessing students' abilities for learning and provide greater focus and specificity of feedback to enhance the quality of instruction. That is, if instructors better understand how their students perceive the educational tasks, instructors can then incorporate this new assessment information to improve instruction.

In the context of this article, basic mathematics items are deconstructed with respect to their intrinsic difficulties, as reflected by common misconceptions (also considered erroneous rules by Tatsuoka et al. 2012) that students report about how they understand what the items are testing. If these misconceptions can be identified on a large scale, then curricula can be developed that are more useful for instructors and ultimately for their students. More specifically, using an improved understanding of their students' misconceptions, instructors should be more effective in remediating students who have learning difficulties. While in the process of developing the prototype for the quantitative literacy assessment instrument for uncovering misconceptions on a large scale with developmental community college students, one instructor reported that students were more engaged in the assessment process. In fact, students appeared to have liked the experience of providing judgments about the test items and reporting their rationales and attempts at solving each of the problems. Students appreciated that the teacher showed an interest in diagnosing student misconceptions (Krishnan and Secolsky 2012).

Determining the prevalence of misconceptions requires that students be presented with correct and/or incorrect solution strategies in order to capture misconceptions for potentially reducing the intrinsic difficulty of items on the assessment. We viewed student selection of incorrect strategies as misconceptions, some common and others not so common. As students become more aware of their misconceptions, they should experience a decline in intrinsic difficulty. This should lead to better problem-solving strategies.

The study of solution strategies has a research history of its own. In particular, for simple to more complex quantitative items, solution strategies assume increasing importance as the complexity of the item increases (Snow and

Lohman 1989). Greeno (1980) claims that in geometry, success in proving theorems depends on systematic planning with goals and subgoals. Snow and Lohman (1989) also point out that successful problem solvers have memorized multiplication facts and other problem-solving strategies. For basic math problems in developmental mathematics courses where syllabi consist mostly of fraction and decimal topics, students with rote memorized math facts will have a better chance at correctly responding to questions, because there will be fewer steps to process cognitively. In other words, the cognitive load will be lower.

Some cognitive psychologists argue that for students to resolve misconceptions, it is often better to let the students directly experience a cognitive conflict by presenting correct new material that runs counter to the misconceptions held by the students (Swan 2001). This cognitive conflict for developmental students is likely to occur if they are first introduced to the misconception they have and are then untaught the misconception. However, the cognitive load introduced through this conflict may be too much for them. They may lose focus, tune out, and become less attentive if forced to confront new correct material unassisted. Interested readers will benefit from Silva and White's (2013) discussion of research related to the psychological problems of developmental mathematics students.

The current research approach is different from traditional assessments of basic skills in two ways. First, solution strategies and misconceptions were derived from actual student reports. Second, the assessment consisted of two parts: solving the actual basic skills items and selecting the correct solution strategies and not selecting incorrect solution strategies (misconceptions). While there has been some research (e.g., Ben-Zvi and Garfield 2004) for identifying misconceptions on a statistics assessment instrument, aside from Sweller's there is generally little research related to the use of solution strategies in the development of a quantitative literacy assessment instrument. The idea is potentially beneficial for instrument development and for identifying misconceptions on larger scales.

In line with the study's objective to identify quantitative literacy misconceptions, we employed logistic regression as a means for characterizing the functioning of the four solution strategies created for each item; all but six of those items consisted of the correct solution and three incorrect solutions, which we classified as misconceptions. For the other six items, there were no correct solution strategies identified in the think-aloud protocols, just misconceptions. In all, 80 such responses were made by each examinee.

By construction, logistic regression produces estimates which are bound between 0 and 1. As such, it is often employed in situations like ours in which the dependent variable is dichotomous: The student is either correct (score = 1) or incorrect (score = 0) in determining whether each strategy is correct or incorrect.

Using logistic regression we estimate the probability that a student correctly characterizes each proposed solution strategy as a function of his/her ability which is measured by the total test score on the 20 item math test. (The total test scores on these items ranged from 0 to17.) After the logistic regression equation was obtained in estimating the probability of each solution strategy, each total test score was substituted into each logistic regression equation. Then for purposes of graphical display, the mean predicted probability for each strategy was obtained for intervals of 3 total test score points for 6 ability intervals: 0-2, 3-5, 6-8, 9-11, 12-14, 15-17. For example, total scores 0, 1, and 2 were each substituted into the logistic regression equation for each solution strategy for each item to obtain these three particular predicted probabilities for each strategy. Then the predicted probabilities were averaged for the 0, 1, and 2 total scores in this interval. The resulting mean predicted probability for each strategy was then plotted and points representing each of the 6 intervals were connected to produce the logistic regression line. In each logistic regression, the dependent variable was the correctness of each student's solution strategy (0 or 1). The independent variable was the student's total test score.

Figure 1 provides an example of results from item 2 on our test. The four lines present the estimated probability that students correctly characterize each strategy in item 2. In the example, the estimated probabilities of properly characterizing strategies 2B, 2C and 2D are positively related to student ability. A positive relationship is generally expected as the ability to correctly sort through strategies should be correlated with the ability to do the problems. The fact that the slope for strategy 2B is steeper than that for 2C and 2D tells us that comprehension of this strategy is particularly associated with higher test scores Such differentiations are related to "discrimination" in classical test theory (CTT). This may suggest that teachers should focus on un-teaching this misconception.
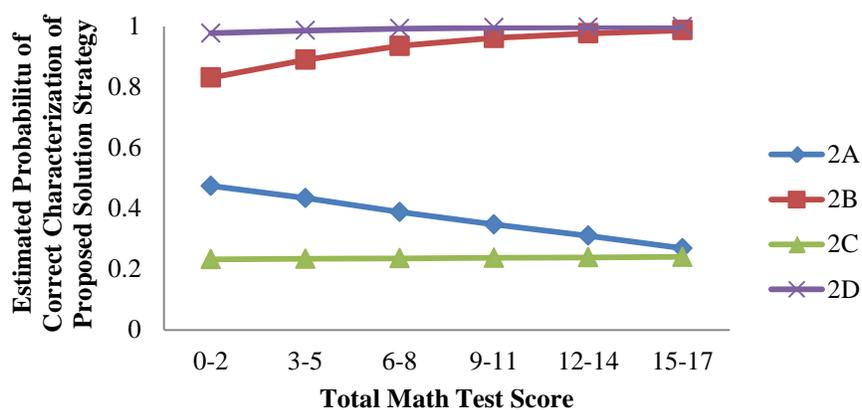

Figure 1. Example analysis of proportion of students correctly characterizing solution strategies for item 2 by students' total scores on the 20-point math test.

Interestingly, ability to characterize strategy 2A is negatively related with total test score. Strategy 2A deals with the misconception of cross-multiplying fractions when the rule is to multiply fractions across. It appears that for this item, higher ability students thought this was a correct strategy and perhaps did not know what was implied by the expression "cross multiply."

The levels of the predicted probabilities shown in Figure 1 are also informative. In CTT, the difficulty of an item is defined as the proportion of examinees responding correctly to a test item. Thus, an easy item would have a value closer to 1.0, while a difficult item would be closer to the value achieved through chance guessing. For example, in Figure 1 the low probabilities of correctly characterizing strategy 2C at all levels of ability means that assessing the quality of this strategy is more difficult than for the others. Interestingly, in this example item and in most of the other problems on the test, accurately characterizing strategies as correct was more difficult than characterizing strategies as misconceptions. This suggests that students disproportionately coded strategies as misconceptions. Thus they were most often correct when the strategy was in fact a misconception and incorrect otherwise. This may partially account for why it was harder for students to accurately characterize correct strategies.

However, item difficulty in CTT is not informative regarding the many ways or reasons why students find items difficult. Understanding the solution strategies underlying student responses to basic mathematics problems could lead to more valid interpretations of test scores than currently offered by traditional CTT item difficulty indices. For test developers and mathematics educators, inquiry into the actual response processes of examinees could offer greater instructional power. Response processes represent an important validation domain, according to the *Standards for Educational and Psychological Testing* (American Educational Research Association, American Psychological Association, and National Council on Measurement in Education 2014). In addition, the new standards emphasize test fairness, and the methodologies employed in the current study show promise for producing new insights on test equity and fairness. As such, this article attempts to demonstrate the benefits of combining the application of technical psychometric procedures with more qualitative approaches toward creation of a quantitative literacy reasoning instrument. This mixed-methods approach contributes to the development of the instrument by the combination of actual test scores and selections of misconceptions from a set of solution strategies. It is an attempt to build into the instrument greater diagnostic information of student understanding of math content and ultimately use this information for enhancing individual and group instruction.

Methods

Participants

Twenty-two students in developmental mathematics classes from a community college in New York volunteered to participate in the think-aloud process at the request of their math skills instructors. Participants were asked to describe their processes for solving five items selected from a 20-item basic math review test given at a community college in New Jersey. From the data collected during the think-aloud process, an instrument was created that was administered to a sample of 238 basic mathematics students in grades 10 to 12 from a high school in northern New Jersey and 209 students in developmental mathematics classes in community colleges in the states of New York and New Jersey. All basic skills classes from the high school were administered the instrument by the mathematics chairperson, and instructors from a number of community colleges in New Jersey voluntarily administered the instrument to their students in intact classrooms. An independent-samples *t*-test on the differences between the means and variances of total score on the instrument for the high school and community college groups revealed differences between the groups. Subsequently, analyses were conducted separately for each group.

A final sample of 30 high school students from the same New Jersey high school (10 treatment and 20 comparison group students) were later administered the new instrument as part of a quasi-experimental pilot study with pre- and post-tests to explore whether this innovative assessment instrument might be effective in pedagogical practice. The students were selected by the mathematics chairperson of the high school, who obtained the cooperation of three instructors.

Procedures

An exploratory sequential mixed-methods design as characterized by Creswell and Plano Clark (2007) was conducted and implemented in four stages for developing the assessment instrument on quantitative literacy and reasoning for developmental skills in mathematics assessment and instruction. Each of the four stages is described as follows.

Stage 1: Selecting Items and Collecting Solution Strategies from Qualitative Think-Aloud Protocols. The first stage consisted of the audio recording of qualitative data from the think-aloud protocols based on the thinking of 22 community college developmental mathematics students. One researcher coded the think-aloud protocol data and rephrased them into a number of so*lution strategies* for each of the 20 fraction and decimal items that made up the test administered to the students. Each of the 20 items on the test was carefully selected from a review test for the final exam for a course in developmental skills.

Think-aloud protocols developed by Ericsson and Simon (1993) were used to ascertain student solution strategies. The protocol consisted of students being asked to speak aloud what they were thinking when responding to items from a review test based on the Accuplacer placement exam. Students who volunteered went into a separate room with one of two researchers who administered the think-aloud protocols to the 22 students. Students were provided a sample question, shown how to share their thoughts, and told that their answers would be recorded for each problem. Each student reported on five items. In each of the five fraction and decimal problems they received, students were asked to solve each problem while continually reporting their thoughts out loud into the tape recorder. In all, the 22 students provided 110 (22 students for five items each) tape-recorded think-aloud protocols.

Stage 2: Instrument Development. Each think-aloud session was transcribed from the tape recordings. From the transcriptions, student solution strategies were extracted so that the key points could be written separately on index cards for each item. The index cards were sorted into 20 piles (one pile for each item). An instrument was then created consisting of both the actual 20 fraction and decimal items and four solution strategies per item derived from the student think-aloud reports. Because students have to answer each question in an open-ended fashion with no partial credit and have to evaluate each solution strategy for its correctness as a strategy for solving each item, the resulting instrument is not a multiple-choice test. That is, for each of the 20 items, students were asked to complete five steps: find the answer to each problem and then evaluate the correctness of each of four solution strategies.

For most items, there were at least three distinct solution strategies per item that emanated from the think-aloud transcripts. For six of the 20 items, no correct strategies were obtained from the 22 students interviewed, just misconceptions. This finding speaks to the potential of our method to identify misconceptions in advance of actual teaching. For these six items, all four solution strategies were incorrect strategies. In four instances where fewer than four distinct solution strategies were identified, two mathematics instructors constructed likely strategies for the assessment instrument. The number of solution strategies was limited to four per item for purposes of limiting testing time. As noted earlier, incorrect solution strategies were treated as mathematical content misconceptions. It should be reiterated that students were instructed to evaluate the correctness of all four solution strategies. We believe it is important to determine each misconception's attractiveness to students, particularly if we intend to develop instructional interventions to directly address incorrect thinking. The assessment instrument consisting of the 20 questions and their four associated solution

strategies was now complete. This assessment instrument appears in Appendix A[1] along with additional information to be described in the Results section.

A note should be made about what constitutes a "strategy" in this context. Student expressions in the think-aloud protocol were given great deference. To those of us with facility in solving math problems, it is readily apparent that some of the "strategies" articulated by students aren't strategies at all. While one could argue that we should have censored the student responses to include only statements which we determined were in fact strategies, we felt this might lead to the loss of valuable pedagogical information. In the protocol we asked students to share the strategies they were using to think through the problem. If the student we engaged in the think-aloud protocol articulated unproductive "strategies," it seems possible—likely even—that other students might be tempted to go in the same direction. After all, the whole point of this exercise is to ferret out the many incorrect strategies students have; we presume faulty thinking abounds. Furthermore, because the characterization of the four strategies is not a multiple-choice exercise, the inclusion of a non-strategic strategy does not alter how students respond to any of the other strategies.

In a similar manner, we intentionally presented strategies as articulated by the students. In some cases this included incorrect grammar. While some might argue that we should have edited the students' strategy statements to "clean up" these errors, such editing runs the risk of altering what the student in question was thinking. The point here is to allow students to reveal how *they* approach these problems. And so we deferred to the ways in which they chose to express those ideas.

Appendix B presents detailed examples of items 5, 8, and 14 illustrating how the information from the raw interviews was extracted and used in the creation of the three sample sets of solution strategies, showing the connection between the think-aloud protocol transcriptions used for the development and presentation of the solution strategies produced for the assessment instrument.

Stage 3: Instrument Administration and Quantitative Analysis of Responses to Questions and Solution Strategy Selections. In this mixed-methods design, stage 3 consisted of quantitative analysis of the responses from the administration of the new assessment instrument. Students were instructed to complete the math test items and to evaluate the correctness of four associated solution strategies.

Stage 4: Preliminary Evidence on the Efficacy of Presenting the Solution Strategies as Part of Pedagogy. It was hypothesized that students would improve in a particular basic skill if they were first untaught the misconceptions. That is, if the misguided steps are revealed first and students are then taught the mathematical concepts from the perspective of the expert, perhaps more

---

[1] Appendices are available in "Additional Files."

successful outcomes would follow. This represents the desired pedagogical design.

To explore the potential power of explicitly teaching/un-teaching to the misconceptions, we performed a small exploratory study. A group of 30 students all completed the 20-question math test at the start of the course as a pre-test. At the end of the term, a treatment group of 10 students, representing one class section of math basic skills, was given the assessment instrument with the solution strategies while the comparison group of 20 other students was administered the same items without the solution strategies. The number of correct responses on the 20-item test administered to the treatment group with the solution strategies was compared to the number of correct responses of the comparison group on the same 20 items without the solution strategies.

This procedure is admittedly only exploratory and does not directly test the usefulness of the teaching practice described above. Specifically, we did not use the strategies assessment to identify misconceptions and then attempt to un-teach them. Rather, we simply cued students to consider alternative strategies as they performed the test. It is plausible that cuing them in this way helped them to consider their strategies in ways that approximated the teaching process we argue for, but other explanations of differential performance between the treatment and control are also possible and are discussed below. In the future, we hope to implement fully the pedagogical process described above to complete a formal test of our hypothesis.

As part of the quantitative analysis of responses, we produced logistic regression plots for characterizing the functioning of the four solution strategies for each item. They are collected in Appendix C.

Results

Development of the Instrument

Table 1 presents the proportion of correct scores in ascending order of item difficulty in CTT for the high school students. The smaller the proportion of correct scores, the more difficult the item. Data are presented separately for the high school and community college groups because of the results of the independent samples t-test ($p = 0.02$) and Levene test of the homogeneity of variance ($p < .0001$) that indicated that the high school and community college groups were statistically different from one another.

Reliability and Validity

Cronbach's (1951) alpha reliability for the responses to the actual 20 items was 0.82. For the recoded 80 solution strategies (1 = correct, 0 = incorrect), alpha was calculated as 0.72, which is deemed acceptable. Validity has to do with the

argument for the plausibility of the interpretation of scores, and solution strategy selections will depend on how the instrument is used. No attempt has been made as yet to correlate these data with other existing variables for these students, such as with actual Accuplacer placement scores. While validation is an ongoing process, the items were constructed to represent the domains for college-level placement testing; as such, the assessment instrument has some degree of content-oriented validity evidence or representativeness of the developmental mathematics domain of skills.

Table 1
Proportion correct scores for problems

		Proportion correct	
Question number[a]	Problem	High school[b]	Community college[c]
15	Divide: 7 1/6 ÷ 1 6/7	0.000	0.020
13	Multiply: 17 4/7 × ¼	0.017	0.134
20	Find percent notation: 5/8	0.034	0.069
14	Divide (write as a mixed numeral): 12 ÷ 1 1/13	0.042	0.060
17	Find decimal notation: 4/15	0.050	0.069
18	Calculate: 1/4 × 1224	0.076	0.112
4	Divide and simplify: 7/2 ÷ 49/4	0.101	0.217
10	Add: 8 1/9 + 7 2/5	0.139	0.201
11	Subtract: 9 2/5 – 5 1/3	0.185	0.222
19	Find percent notation for: 0.372	0.214	0.213
12	Subtract (write a mixed numeral for the answer): 27 – 22 ½	0.227	0.242
6	Add and simplify: 7/8 + 7/8	0.273	0.248
7	Add and simplify: 7/9 + 5/6	0.277	0.282
9	Add (write the answer as a mixed numeral): 6 5/6 + 2 5/6	0.307	0.257
2	Multiply and simplify: 2/5 × 35	0.336	0.394
5	Divide and simplify: 7/4 ÷ 7	0.340	0.374
3	Multiply and simplify: 3/10 × 43/100	0.345	0.376
16	Round to the nearest tenth: 7.8493	0.357	0.340
8	Subtract and simplify: 7/10 – 13/25	0.391	0.380
1	Simplify: 9/15	0.895	0.848

[a] Questions are sorted in order of increasing difficulty for high school participants.
[b] $N = 238$.
[c] $N = 209$.

Exploratory Analysis

Analysis of covariance (ANCOVA) was conducted on the mean total actual 20-item post-test using as covariates pretest scores, previous year's high school final math grades and a dummy variable noting whether the groups were presented with (treatment group) or not presented with (comparison group) the solution strategies (correct or misguided) as part of the post-test assessment (Table 2).

According to the analysis, participation in the treatment/control group has far greater explanatory power than either the pretest score or previous year's GPA.

Table 2.
ANCOVA Analysis of Post-Test Scores.

Source	Df	Mean square	F	p
Pretest score	1	7.79	1.26	0.27
Last year's GPA	1	0.23	0.04	0.85
Strategies present	1	63.25	10.27	0.0033
Error	29	6.16		

Table 3 shows the difference in adjusted mean scores after performing the analysis of covariance with the treatment group having significantly higher mean scores than the comparison group. Those asked to evaluate alternative strategies while taking the post-test scored more than 3 points (or 60 percent) higher ($p = .003$). We believe that this evidence is suggestive of the power of having students confront misconceptions. If we are right, then instructors in developmental math may be able to substantially increase student performance if they begin by engaging students in think-aloud protocols or other methods to unearth misconceptions to be untaught.

Table 3.
Adjusted Mean Scores Out of 20 Items, Controlling for Other Covariates.

Condition	Mean	p
Assessed with instrument	8.13	0.0033
Traditional	5.01	

Of course, in this limited and exploratory analysis we did not have time to collect student strategies and then alter the curriculum. Thus, alternative explanations for the results in Table 3 must be acknowledged. It is possible that presenting students with correct strategies (mixed with incorrect ones) on 12 of the 20 items was enough to prompt recall. On the other hand, students may have simply implemented a randomly chosen strategy. (If the mean score in the control group was 5 correct, randomly choosing one from four listed strategies on the remaining 15 problems could lead to an increase of more than 3 points if students properly implemented the correct strategy when it was selected.) While we can imagine such alternative explanations, the inherent logic behind targeting instruction to unlearn misconceptions is strong enough that we look forward to testing formally the effects of this pedagogical practice in future research.

Discussion and Conclusion

The process described and instrument developed for assessing quantitative literacy reasoning has the potential for enabling teachers and students to address

more directly misconceptions in basic mathematics content earlier and more effectively than with traditional pedagogy. With this instrument even as a prototype we believe instructors can obtain clearer ideas of the conceptions and misconceptions students have for solving problems. We believe students can benefit, particularly in lower ranges of the ability spectrum.

The ability to distinguish correct from incorrect solution strategies (misconceptions) goes further into developing mathematics ability than merely solving the actual problems. It requires understanding what is wrong with a given solution strategy. This understanding will benefit those in need of basic or developmental mathematics.

Not only is the assessment instrument useful for classroom purposes and departmental unit exams, but it also may benefit curriculum developers. Knowing what students know and how they commonly misconceive content can foster the development of more appropriate curriculum materials for learners. The proportions of students selecting each solution strategy inform both pedagogical practice and curriculum development in basic mathematics. The logistic plots can be valuable in assisting instructors on focusing their instructional time more efficiently.

Finally, preliminary data collected indicate that students having access to the set of correct and incorrect solution strategies for items in a test may enhance their performance on the test itself. This does not demonstrate that students would score higher on the actual test when assessed in basic skills if untaught the misconceptions first before being taught correct strategies from the perspective of the expert. However, it leaves the question yet to be determined in a controlled experimental study.

Several limitations were present in the development of the instrument. First, only one researcher coded and constructed the majority of solution strategies from the think-aloud protocol data. Second, the group of 22 students from which the solution strategies were extracted may not reflect other potential misconceptions or erroneous rules. Third, for the large-scale collection of solution strategy selections, the point in the instructional unit of study where data were collected differed somewhat across the community colleges. All that is known is that for the developmental community college students, data were collected midway in the semester.

It is important to realize from the data collected that many students are deficient in basic math as well as other developmental subjects in the United States and around the world. While strategies for increasing accountability for teachers exist, and restructuring of curricula provide different types of opportunities for success for students, a greater focus on instructional innovation is also necessary. The proposed process and assessment instrument moves the field in a different direction—not to one of enforcement of accountability

standards but rather to one of increased basic learning and educational achievement. By listening to our students as they solve problems we should be able to learn not only what they struggle with but also why. Teaching to these misconceptions rather than to the test may be a better way to help students navigate their way through developmental mathematics.

Acknowledgments

References

American Educational Research Association, American Psychological Association, and National Council on Measurement in Education. 2014. *Standards for educational and psychological testing.* Washington, DC: Author.

Askew, M., and D. Wiliam. 1995. *Ofsted reviews of research: Recent research in mathematics education* 5–16. London: HMSO.

Ben-Zvi, D., and J. Garfield, eds. 2004. *The challenge of developing statistical literacy, reasoning and thinking*. Dordrecht, Netherlands: Kluwer Academic. http://dx.doi.org/10.1007/1-4020-2278-6

Bjork, R. A. 1994. Memory and metamemory considerations in the training of human beings. In *Metacogniton: Knowing about knowing,* ed. J. Metcalfe and A. Shimamaura, 188-205. Cambridge, MA: MIT Press.

Creswell, J. W., and V. L. Plano Clark. 2007. *Designing and conducting mixed methods research.* Thousand Oaks, CA: Sage.

Cronbach, L. J. 1951. Coefficient alpha and the internal structure of tests. *Psychometrika* 16: 297–334. http://dx.doi.org/10.1007/BF02310555

Cross, K., and T. A. Angelo. 1993. *Classroom assessment techniques: A handbook for college teachers.* San Francisco: Jossey-Bass.

Darling-Hammond, L. May 2, 2007. Evaluating "No Child Left Behind." *The Nation*. http://www.thenation.com/article/evaluating-no-child-left-behind

Drezek-McConnell, K., and P. E. Doolittle. 2012. Classroom-level assessment: Aligning pedagogical practices to enhance student learning. In *Handbook on measurement, assessment, and evaluation in higher education*, ed. C. Secolsky and D. B. Denison, 15–30. New York: Routledge.

Ericsson, K. A., and H. A. Simon. 1993. *Protocol analysis*: *Verbal reports as data* (revised edition). Cambridge, MA: MIT Press.

Feltovich, P. J., R. J. Spiro, and R. L. Coulson. 1993. Learning, teaching and testing for complex conceptual understanding. In *Test theory for a new generation of tests,* ed. N. Fredericksen, R. J. Mislevy, and I. Bejar, 181–217. Hillsdale, NJ: Erlbaum.

Greeno, J. G. 1980. Some examples of cognitive task analysis with instructional implications. In *Aptitude, learning, and instruction: Vol.2,* ed. R. E. Snow, P. A. Frederico, and W. E. Montague, 1–21. Hillsdale, NJ: Erlbaum.

Krishnan, S., and C. Secolsky. 2012. *Developmental mathematics and innovative instruction*. Roundtable. Forum for the Association for Institutional Research, New Orleans.

Lacireno-Paquet, N., C. Morgan, and D. Mello. 2014. How states use student, learning objectives in teacher evaluation systems: A review of state websites. Washington, DC: U.S. Department of Education.

Provezis, S. October 2010. Regional accreditation and student learning outcomes: Mapping the territory (NILOA Occasional Paper No. 6). University of Illinois and Indiana University, National Institute of Learning Outcomes Assessment.

Schray, V. 2006. Assuring quality in higher education: Key issues and questions for changing accreditation in the United States. Washington, DC: Secretary of Education's Commission on the Future of Higher Education.

Silva, E. and T. White. 2013. *Pathways to improvement: Psychological strategies to help college students master developmental math*. Palo Alto, CA: Carnegie Foundation for the Advancement of Teaching.

Snow, R. E., and D. F. Lohman. 1989. Implications of cognitive psychology for educational measurement. In *Educational measurement,* 3rd ed., ed. R. L. Linn, 263-331. New York: American Council on Education, Macmillan.

Swan, M. 2001. Dealing with misconceptions in mathematics. In *Issues in mathematics teaching*, ed. P. Gates, 147–165. London: Routledge-Falmer.

Sweller, J. 1988. Cognitive load during problem solving: Effects on learning. *Cognitive science* 12(2): 257–285. http://dx.doi.org/10.1207/s15516709cog1202_4

———. 1994. Cognitive load theory, learning difficulty, and instructional design. *Learning and instruction* 4(4): 295–312. http://dx.doi.org/10.1016/0959-4752(94)90003-5

———. 2010. Element interactivity and intrinsic, extraneous and germane cognitive load. *Educational and psychological review* 22(2): 123–38. http://dx.doi.org/10.1007/s10648-010-9128-5

Tatsuoka, K., C. Tatsuoka, A. Kelly, and M. Dean. 2012. Cognitive diagnostic method: Rule space, Q-matrix theory and applications. In *Handbook on measurement, assessment, and evaluation in higher education*, ed. C. Secolsky, and D.B. Denison. 438–451, New York: Routledge.

U.S. Department of Education. 2006. *A test of leadership: Charting the future of U.S. higher education*. Washington, DC: Author.