Hammill Institute on Disabilities

The Journal of Special Education
45(2) 67–76

DOI: 10.1177/0022466909351579
http://journalofspecialeducation.sagepub.com
SAGE

# Detecting Cognitive Change in the Math Skills of Low-Achieving Adolescents

**Sun-Joo Cho[1], Brian A. Bottge[2], Allan S. Cohen[3], and Seock-Ho Kim[3]**

## Abstract

Current methods for detecting growth of students' problem-solving skills in math focus mainly on analyzing changes in test scores. Score-level analysis, however, may fail to reflect subtle changes that might be evident at the item level. This article demonstrates a method for studying item-level changes using data from a multiwave experiment with a teaching method called enhanced anchored instruction (EAI). The analysis combines a mixture Rasch model for detecting individual differences within latent groups with a latent transition analysis model for tracking changes in latent group membership over the course of EAI. The analysis clearly indicates the effects of EAI and how they differ for members of each latent class. Comparisons are provided with a standard analysis of changes in test scores. Implications of the new approach are discussed for detecting subtle transformations in the math performance of students across a range of ability levels.

## Keywords

mixture item response theory model, latent transition model, math, assessment, learning disabilities, enhanced anchored instruction

The National Mathematics Advisory Panel (2008) recommended expanding the ways that skills are measured because assessment drives instructional decision making. According to the panel, more collaboration is needed between psychometricians, who have specialized knowledge in data analysis models, and teaching professionals, who are most familiar with the constructs that the tests are intended to measure. The result, it is hoped, will be new kinds of assessments that are of high quality, more sensitive to new forms of instruction, and more helpful to teachers in customizing their instruction for improved student performance (Pellegrino, Chudowsky, & Glaser, 2001).

Typical methods for detecting instructional effects are at the score level, and modeling changes in student achievement are characterized as observed or latent variables on total test scores. For example, growth modeling, such as with a hierarchical linear model, is based on observed scores. No latent variable is modeled to incorporate measurement error, although random effects in the model are arbitrarily assumed to be normally distributed in the same way as is typically assumed for ability in item response theory (IRT) models. Latent growth curve approaches (Willett & Sayer, 1994) are also modeled with observed scores, but growth is characterized in terms of an initial state and growth rate as the latent variables. IRT-based longitudinal models (Andersen, 1985; Embretson, 1991; Fischer, 1976, 1989, 1995) treat ability as a latent variable. This approach can be useful for assessing overall effects of treatment or change. Although such models permit quantitative differences to be studied, they do not provide information on how students' response processes change over time.

In addition to examining total test scores, important new information can be gleaned by examining students' responses to individual test questions. This item-level information can be quantitative, qualitative, or both. In this article, we use a new method that combines an existing data analysis method, latent transition analysis (LTA; Collins & Wugalter, 1992), with a mixture IRT model, a mixture Rasch model (MRM; Rost, 1990, 1997), to detect possible changes in growth patterns of students across a range of ability levels. The LTA-MRM method (Cho, Cohen, Kim, & Bottge, 2010) is based on previous research that has demonstrated how mixture IRT models can identify latent groups of students who differ in the kinds of cognitive strategies used to answer test questions (Embretson & Reise, 2000; Rost, 1990, 1997). LTA-MRM tracks changes in the use of cognitive strategies, the level of success on each item, and the latent class membership change due to instructional intervention.

[1]Vanderbilt University, Nashville, TN, USA
[2]University of Kentucky, Lexington, KY, USA
[3]University of Georgia, Athens, GA, USA

**Corresponding Author:**
Sun-Joo Cho, Vanderbilt University, Peabody H213a, 230 Appleton Place, Nashville, TN 37203, USA
E-mail: sj.cho@vanderbilt.edu

The purpose of this study is to demonstrate how the LTA-MRM can be used to reveal changes in growth patterns of students who participated in previous research with an instructional approach called *enhanced anchored instruction* (EAI; Bottge, Rueda, Serlin, Hung, & Kwon, 2007). The theoretical and practical issues examined in the reevaluation of the data are particularly important for two reasons. First, it is hoped that a new model of assessment will emerge that results in tighter connections between teaching and measurement. This new model, which includes format, analysis, and feedback, will demonstrate the reciprocal benefits of teaching for cognitive change and tracking the characteristics of that change (Pellegrino et al., 2001). Second, we hope that these practices can assist special educators in developing and refining response to intervention (RTI) methods as part of a procedure for identifying students with learning disabilities (LD; Fuchs, 2003; Fuchs & Deshler, 2007; Fuchs & Fuchs, 2006).

In this article, we summarize the 2007 host study that was aimed at improving the problem-solving skills of students with and without LD. Next, we report the use of LTA-MRM to reanalyze the data and examine possible changes in item-level performance. Finally, we suggest instructional (i.e., EAI) and measurement (i.e., LTA-MRM) implications based on the results.

## Method

### Host Study

*Participants*. Complete data sets of test scores from students ($N$ = 109) in six seventh-grade mathematics classrooms were drawn from the host study for use in this reanalysis. Four classes of students were typically achieving, one class was high achieving (pre-algebra), and one inclusive class consisted of 13 students with an Individualized Education Program (IEP; 12 students with LD) and 13 typically achieving students. Two math teachers (MT1, MT2) each taught three 90-minute classes per day. MT1 taught a class of typically achieving students, the pre-algebra class, and the inclusive class. To be included in the pre-algebra class, students achieved at or above the 90th percentile on the state math test, posted high scores on teacher-made tests, and demonstrated excellence on daily assignments. In the inclusive class, students with IEPs qualified for special education services according to criteria set by the Wisconsin Department of Public Instruction (2006). That is, students with LD had a significant discrepancy equal to or greater than 1.75 standard errors of the estimate below expected achievement based on individually administered, standardized achievement and ability tests. They received special education services an average of 635 minutes per week (range = 300–920). Mean national percentile (NP) ranks on the *Iowa Tests of Basic Skills* (ITBS; Hoover, Dunbar, & Frisbie, 2001) administered as pretests immediately prior to intervention confirmed these students' low math achievement: NP = 15 in math computation and NP = 14 in problem solving.

Based on pretest results of the ITBS, MT1's three classes differed in computation and in problem solving. Pre-algebra students outscored students in the typical classes in computation and problem solving, and students in the typical class scored higher than students in the inclusive class in both computation and problem solving. MT2's typical classes did not differ in computation or in problem solving, nor were there differences between the four typical classes (one of MT1 and three of MT2) in computation or in problem solving.

*EAI*. Between Time Point 2 (Pretest 2) and Time Point 3 (posttest) teachers taught two EAI modules to students in all six classes. EAI is a form of anchored instruction, a teaching approach made popular by the Cognition and Technology Group at Vanderbilt (1990, 1997). EAI consists of one overarching problem comprised of several subproblems presented in a multimedia format. After this problem is solved, students work in small groups on an applied, hands-on problem that requires the same skill set as the multimedia-based problem. EAI shares common core characteristics with problem-based learning (Barrows, 1996), which is often used in medical education (Gijbels, Dochy, Van den Bossche, & Segers, 2005). For example, instructors lead discussions with interesting questions, students work together in small groups to develop and test solutions to the problems, and instructors provide explicit skill instruction as needed. Learning scaffolds embedded in the EAI problems provide students with multiple opportunities to practice their skills in varied contexts, which is important for skills and concepts transfer (Brown, Collins, & Duguid, 1989; Greeno & the Middle School Mathematics Through Applications Project Group, 1998).

The multimedia-based problem used in the host study is called "Fraction of the Cost" (FOC) and stars three students from a middle school. One actor is a girl with Down syndrome whose parents encouraged her participation and helped her rehearse her lines. The video was filmed at a local skateboarding store and in a computer room, garage, and backyard of a local home. The teens study a schematic plan of a skateboard ramp and then discuss how they can afford to build a ramp with the materials and money they have available. To solve the problem, students need to (a) calculate the percentage of money in a savings account and sales tax on a purchase, (b) read a tape measure, (c) convert feet to inches, (d) decipher building plans, (e) construct a table of materials, (f) compute mixed fractions, (g) estimate and compute combinations, and (h) calculate total cost. Several instructional modules supported teaching, learning, and assessment (see Figures 1 and 2).

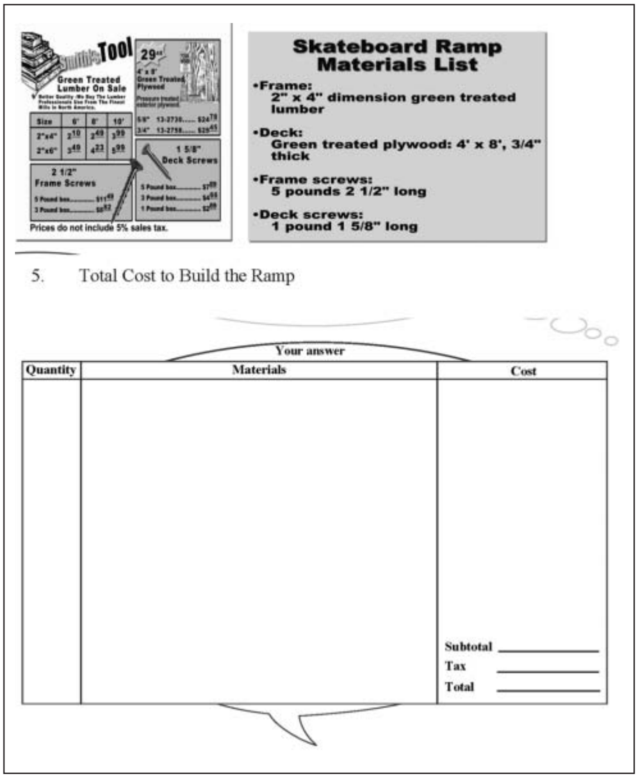


**Figure 1.** Example of a complex item.
Source: Adapted from *Mathematics Teaching in the Middle School*, copyright May 2009 by the National Council of Teachers of Mathematics. All rights reserved.

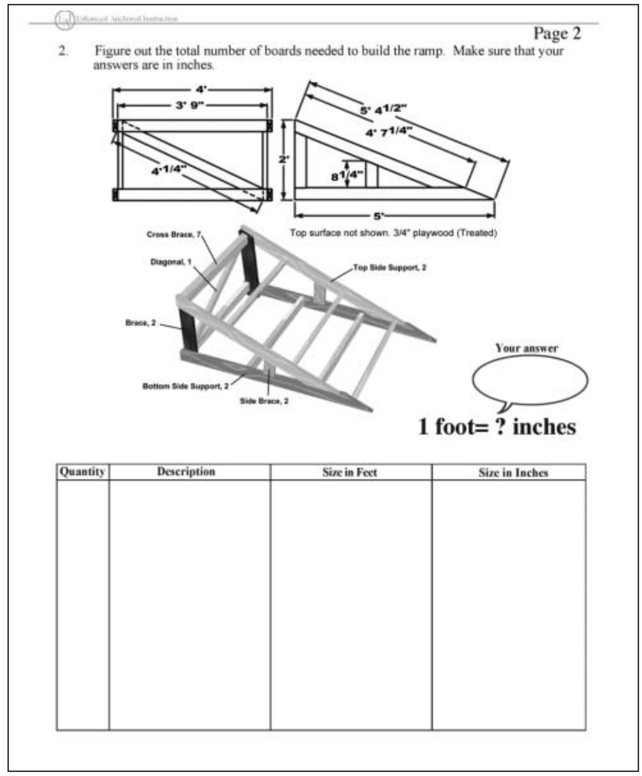


**Figure 2.** Example of a single-skill item.
Source: Adapted from *Mathematics Teaching in the Middle School*, copyright May 2009 by the National Council of Teachers of Mathematics. All rights reserved.

A hands-on problem about a hovercraft requires students to apply what they learned from FOC by planning, drawing, and constructing a "rollover cage" for a hovercraft out of PVC pipe. Each student draws a schematic plan showing several views of the cage and builds a scale model out of plastic straws. Then, students vote to select the three designs they want to make. The teacher divides the class into three groups, and each group plans how they can make the cage in the most economical way, which involves cutting the 10-foot lengths of pipe in such a way as to waste as little pipe as possible. Once the teacher approves the plans, students work on measuring, cutting, and assembling. When the cages are complete, they lift them onto a 4 × 4-foot plywood platform (i.e., hovercraft). A leaf blower inserted into a hole in the plywood powers the hovercraft, inflating the plastic attached to its underside and elevating the hovercraft slightly above the floor. Students ride on the hovercrafts in relay races on the last day of the project.

*Instrumentation*. The *Fraction of the Cost Challenge* (FOCC) consists of constructed response items designed to measure the skills and concepts taught with FOC. The content of the items is aligned with the National Council of Teachers of Mathematics (2000) standards recommended for middle school students (Measurement, Problem Solving, Representation). Versions of the test had been used successfully in previous studies (e.g., Bottge, Heinrichs, Mehta, & Hung, 2002; Bottge et al., 2004). Sets of problems were weighted according to their complexity and the contribution they were expected to make in solving the main problem. Within each set, items were awarded full or partial credit, which made it possible to analyze student work at both the item and concept levels. The concurrent validity correlation coefficient of the FOCC based on pretest scores of the ITBS was .59, which seems acceptable given that the range of the mathematics concepts sample of the FOCC was more restricted than that of the ITBS. The internal consistency estimate (α) was .92. Interrater reliability on 20% of the protocols selected from each of the test administrations was 94% (range = 90%–97%) and was computed by dividing the number of agreements by the total number of agreements and disagreements and multiplying by 100 (Sulzer-Azaroff & Mayer, 1977).

The FOCC measured students' ability to interpret a three-dimensional schematic plan, measure lengths of building materials in feet and inches, estimate and compute combinations using whole numbers and fractions, interpret and record data in tables, and calculate total cost including sales tax. The test emphasizes real-world knowledge such as reading a bank account statement, comparing treated and untreated lumber, selecting appropriate lengths and quantities of screws for specific applications, and developing a materials list with a cost estimate. Students use rulers to measure the lengths of boards that are shown in a schematic plan of a building project. Students are to record the lengths to the nearest 1/8 inch in feet and inches or in inches alone. To solve this problem, students must first be able to "read" the schematic plan, identify the number of boards of each length, and then figure out the most economical way to use the available wood. Then, they use these combinations to indicate where to saw the boards.

*Research design*. A multiwave, multimeasure design was used to determine the effects of EAI on students' math performance. Data taken from the FOCC tests at three time points (i.e., double pretests, one posttest) over the course of the school year were analyzed in this study. Fourteen weeks separated the first pretest (Time 1) and second pretest (Time 2), when teachers followed their standard curriculum. The standard curriculum between pretests consisted of units related to geometry and proportional reasoning in *Connected Mathematics* (Dale Seymour Publications, 2004). Teachers organized their instruction in ways that were similar to methods used during EAI. That is, teachers reviewed the previous day's work, taught new concepts and skills explicitly, and then asked students to solve problems in small groups. Math content focused primarily on projects, such as figuring out the cost of an advertisement based on the proportion of the page it covered. The FOCC posttest (Time 3) was administered again immediately following EAI to assess what students had learned between Time 2 and Time 3.

## Assessing Cognitive Changes of EAI

*LTA with an MRM*. An LTA with an MRM measurement model (Cho et al., 2010) was used to assess the effects of EAI in the multiwave experiment. The LTA-MRM determines whether sets of response patterns may be due to changes in differential use of strategies for answering test items. In this particular case, the LTA-MRM was used to detect latent classes in the achievement data from the EAI intervention and determine probabilities of transition paths between latent classes.

The transition pattern in an LTA-MRM describes a particular transition path among the different latent classes and is coded accordingly. For example, suppose there are two test administration periods, Time 1 and Time 2, and a student who is tested at Time 1 is determined to be in Latent Class 1 and at Time 2 is found to be in Latent Class 2. Then the transition pattern would be described as Transition Pattern 1 2. Similarly, if there were a third test administration at Time 3, and a student was determined to be in Class 1 at Time 1, in Class 2 at Time 2, and in Class 2 at Time 3, the transition pattern would be represented as 1 2 2. Each of these transition patterns characterizes a different RTI, and students with the same transition pattern all have the same RTI. In this study, we use these transitions to describe different patterns of change over the course of the intervention.

*Fitting the MRM at each time period*. The LTA-MRM analysis first explored how an MRM fit the data at each time period. This was done using the computer program WINMIRA (von Davier, 2001). The usual procedure in selecting the number of latent classes is to fit separate models to the data where each of the candidate models fits a different number of latent classes. Then, an information index, such as the Bayesian information coefficient (BIC), helps determine which of these models fits best. BIC has been shown to perform well for selecting the correct number of latent classes in simulated data for the MRM (Li, Cohen, Kim, & Cho, 2009). The BIC was used in this study as a statistical index to aid in model selection, and only those FOC test items that required more than one math skill or operation were analyzed. These items were more complex than items that required only a single mathematics skill. Table 1 presents a matrix indicating the mathematics skills that were required to solve each item. (A "1" in Table 1 indicates a skill is required to solve the item, and a "0" indicates a skill is not required.) As can be seen in Table 1, Items 3, 4, 5, 6, 7, 8, 10, and 11 required two or more different skills for a correct answer, so they were considered complex items. Items 1, 2, 9, 12, 13, and 14 were not considered complex because they required only a single skill. These latter items were examined as possible anchors for the metric across latent classes. Anchoring the metric ensures scale comparability among parameters in the two latent classes.

The equality of the item parameters across the latent classes was tested to determine whether the item could be used as an anchor across the two classes. To serve as an anchor item, an item needs to be the same difficulty (i.e., class invariant) in each latent class. Determining class invariance proceeded by constraining the item difficulties separately, that is, for one item at a time over the two latent classes. A likelihood ratio test was used to compare the likelihood of the constrained model to the likelihood of a model in which all items were fully relaxed. A relaxed item is one that is allowed to be different in the two latent classes. In this analysis, the constrained model consisted of a one-item-constrained model in which only the studied item was constrained to be equal in each of the latent classes. The likelihood ratio test compared the value of $-2 \times \log$ likelihood for the one-item-constrained model and $-2 \times \log$ likelihood for the model in which all items were fully relaxed. In this way, each item was examined separately based on data from the first time point to determine whether the parameters of that item were class invariant, that is, whether they were the same in each of the latent classes. No change in the value of $-2 \times \log$ likelihoods was observed between

**Table 1.** Cognitive Skills Required for Solving Each Item

| | Cognitive skill | | |
|---|---|---|---|
| Item number | Number and operations | Measurement | Representation |
| 1 | 1 | 0 | 0 |
| 2 | 1 | 0 | 0 |
| 3 | 1 | 1 | 0 |
| 4 | 1 | 1 | 0 |
| 5 | 1 | 1 | 0 |
| 6 | 1 | 1 | 0 |
| 7 | 1 | 1 | 0 |
| 8 | 1 | 1 | 0 |
| 9 | 0 | 1 | 0 |
| 10 | 1 | 1 | 1 |
| 11 | 1 | 1 | 1 |
| 12 | 1 | 0 | 0 |
| 13 | 1 | 0 | 0 |
| 14 | 1 | 0 | 0 |

**Table 2.** Item Difficulty Estimates

| | Class 1 | | Class 2 | |
|---|---|---|---|---|
| Item | Difficulty estimate | *SE* | Difficulty estimate | *SE* |
| 1 | −0.774 | 0.39 | −0.805 | 0.26 |
| 2 | −1.064 | 0.49 | −1.064 | 0.49 |
| 3 | 3.637 | 2.34 | −0.244 | 0.10 |
| 4 | 3.710 | 2.34 | −1.547 | 0.42 |
| 5 | 4.677 | 2.74 | −2.713 | 0.53 |
| 6 | 4.371 | 2.29 | 0.197 | 0.09 |
| 7 | 2.884 | 1.36 | −2.315 | 1.40 |
| 8 | 2.307 | 1.45 | −1.817 | 0.91 |
| 9 | −1.091 | 0.46 | −1.474 | 0.49 |
| 10 | 0.422 | 0.17 | −0.195 | 0.09 |
| 11 | −0.103 | 0.04 | 0.895 | 0.37 |
| 12 | 0.480 | 0.16 | 0.480 | 0.16 |
| 13 | 1.338 | 0.70 | 1.338 | 0.70 |
| 14 | 0.648 | 0.22 | 0.862 | 0.20 |

the fully relaxed model and the one-item-at-a-time restricted models for Items 2, 12, and 13. This was interpreted to mean that these three items were class invariant because they were found to have the same respective difficulties in Class 1 and Class 2. Based on this result, these three items were used as anchor items for the LTA-MRM analysis. In addition, the mean of ability for the first time point and the first class was set to 0 for model identification. Joint estimation was then used to estimate the remaining item difficulties over the three time periods using the computer program Mplus (Muthén & Muthén, 2006).

The LTA-MRM was first fit by setting all covariances as unconstrained and, therefore, estimated freely. Because data were sparse in some of the response patterns, there were some difficulties in estimating covariance terms for the ability structure. Transition probabilities estimated between time periods were quite small and appeared to be poorly estimated. As a result, the following ability covariance terms were fixed at 0: between Class 2 at Time 1 and Class 1 at Time 2, between Class 1 at Time 2 and Class 1 at Time 3, and between Class 2 at Time 2 and Class 1 at Time 3.

First- and second-order LTA-MRM analyses were then estimated. A first-order LTA-MRM is one in which the transition probabilities are estimated between adjacent time points. A second-order LTA-MRM is one in which the transition probabilities are estimated over the two previous time points. In this study, a first-order LTA-MRM solution was selected because BIC suggested it to be a better fit to the data. LTA-MRM results for 50 starting values in Mplus indicated that the highest log likelihood was replicated in five final stages of solutions, indicating that a local solution probably was not reached and the model was identified for the first-order joint estimation.

## Results

### *Characteristics of Latent Classes*

The BIC indices indicated a two-group model was the best fit at each of the three time points. Characteristics of the two latent classes (i.e., Class 1 and Class 2) were next investigated with respect to item difficulty and proportions of group membership. The item difficulties for the LTA-MRM were estimated using joint estimation over the three time periods based on the assumption of invariance of item parameters across the three time periods. The item difficulty estimates are shown in Table 2.

Figure 3 represents plots of the item difficulties for each of the two latent classes (see Table 2). The plot shows that Item 1 and Items 3 through 10 were clearly more difficult for members of Class 1, and Items 11 and 14 were more difficult for students in Class 2. As indicated in Table 1, Items 3 through 8, 10, and 11 required two or more different skills for a correct answer and were considered to be complex items. Items 3 through 8 asked students to first interpret a schematic plan, then list the number and lengths of wood required for building the skateboard ramp, and finally convert the measurements from feet and inches to inches (see Figure 1). Students in Class 2 were more likely to solve the complex items (i.e., Items 3 through 8, 10, and 11) than students in Class 1. For Items 10 and 11, students had to figure out and show how to cut 2 × 4s to waste as little wood as possible. This task is not straightforward because it requires students to use the most economical combinations of wood from the garage. Students needed to be able to read a tape measure to measure the wood (i.e., measurement), figure out how this wood can be used based on their interpretation of the schematic plan (i.e., representation), and then compute

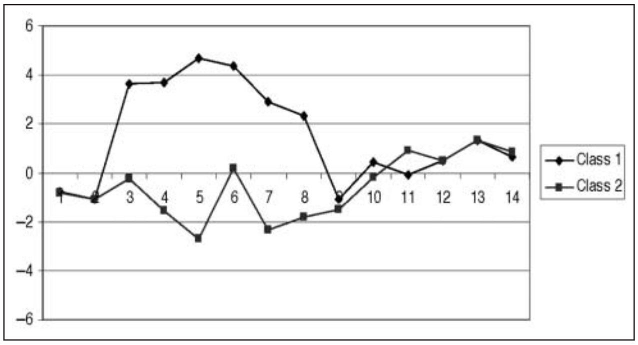


**Figure 3.** Profiles of item difficulties.

the combinations (i.e., numbers and operations). Items 1 and 2 asked students to figure out how much money one of the friends should contribute to the skateboard ramp project. In the video, one boy said that he can spend 10% of the $210 in savings. Looking at a bank statement, his friend states that he must keep $50 of his lawn mowing money in the bank. Items 12 to 14 were considered single-skill items and not mathematically complex, although answers to the problems depended on a correct interpretation of the overall problem posed in FOC and their work on previous items on the test. For example, Item 14 asked students to figure out the total cost of building the ramp, which requires students to read the store advertisement and add the dollar amounts of the items (e.g., 2 × 4s, plywood, screws; see Figure 2). However, students can calculate the cost correctly only by interpreting the schematic plans of the ramp, figuring out the economical use of wood, and calculating how much new wood they need to purchase in previous items.

Membership in Class 2 increased dramatically following EAI instruction. All of the 17 low-achieving students and the 8 students with LD who had been in Class 1 at Time 1 were subsequently classified in Class 2 at Time 3 (recall that Time 3 was the posttest). This transition from Class 1 to Class 2 clearly shows that EAI was effective in teaching the low-achieving students and the students with LD to solve complex mathematics problems that they could not solve before the intervention. In addition, a $\chi^2$ on latent class-by-achievement level and latent class-by-LD status indicated that both achievement level and LD status were related to latent class membership for the first two time periods (i.e., before FOC instruction) for achievement level at Time 1, $\chi^2(2) = 6.137$, $p = .046$; achievement level at Time 2, $\chi^2(2) = 13.918$, $p = .001$; LD status at Time 1, $\chi^2(1) = 3.684$, $p = .045$; and LD status at Time 2, $\chi^2(1) = 10.503$, $p = .001$. However, no association for these variables was found following FOC instruction for achievement level, $\chi^2(2) = 0.360$, $p = .165$, or for LD status, $\chi^2(1) = 0.991$, $p = .320$. Also, gender was not related to class membership at Time 1, $\chi^2(1) = 0.063$, $p = .082$; Time 2, $\chi^2(1) = 0.011$, $p = .916$; and Time 3, $\chi^2(1) = 2.968$, $p = .085$.

**Table 3.** Counts and Proportions for Each Transition Pattern

| Pattern | Without disabilities | With disabilities | Count | Proportion |
|---|---|---|---|---|
| 111 | 5 | 0 | 5 | .046 |
| 112 | 22 | 8 | 30 | .275 |
| 121 | 2 | 0 | 2 | .018 |
| 122 | 27 | 0 | 27 | .248 |
| 211 | 1 | 0 | 1 | .009 |
| 212 | 6 | 0 | 6 | .055 |
| 221 | 2 | 0 | 2 | .018 |
| 222 | 35 | 1 | 36 | .330 |

**Table 4.** Transition Probabilities for Each Latent Class

| Time 1 | Time 2 latent class | |
|---|---|---|
| Latent class | 1 | 2 |
| 1 | .550 | .450 |
| 2 | .166 | .834 |
| **Time 2** | **Time 3 latent class** | |
| Latent class | 1 | 2 |
| 1 | .174 | .826 |
| 2 | .085 | .915 |

## *Latent Transition Probabilities*

The proportions of students in each of the changing patterns, transition probabilities, and quantitative differences in math ability over time for each pattern are provided in Tables 3 and 4. These transition patterns describe the movement from one latent class to the other latent class over the course of the study. Of particular interest are the patterns of students whose ability decreased, stayed the same, or improved for solving the more complex mathematics problems. As a reminder, transition patterns indicate the sequences in which membership of students in latent classes changed or remained the same over the three time points (i.e., pretest, pretest, and posttest) and are another indicator of the change in learning over time. The transition probability indicates the proportion of students of a given latent class at one time point who remained in that class or moved into the other latent class at the next time point. The three most frequently observed transition patterns for all students were 112 (30 students), 122 (27 students), and 222 (36 students). Eight of the 9 students with LD had the transition pattern 112, and 1 student with LD student had pattern 222.

A total of 92% of the students who were in Class 1 at Time 2 (Pretest 2) moved into Class 2 at Time 3 (posttest). This movement of students from Class 1 to Class 2 is evidence supporting the effectiveness of the EAI intervention. However, of the students who were in Class 1 at Pretest 1, 45% were subsequently classified into Class 2 at Pretest 2.

This is an interesting result as no FOC intervention occurred until after Pretest 2. It is possible that this transition probability reflects a memory effect because the same performance tasks were used across all three tests. Another possibility is that the math curriculum taught during the interim between pretests helped students better understand some items on the FOC test.

### Mathematics Performance of Students Within Latent Classes Over Time

On the FOC test, there were 64 students in Class 1 at Time 1 ($M$ = 0, $SD$ = 0.93), 42 students in Class 1 at Time 2 ($M$ = 0.45, $SD$ = 0.78), and 10 students in Class 1 at Time 3 ($M$ = 0.95, $SD$ = 0.47). As a reminder, the mean ability for the first time point and the first class was set at 0 for model identification. In Class 2, 45 students were at Time 1 ($M$ = 0.42, $SD$ = 0.73), 67 students were at Time 2 ($M$ = 1.01, $SD$ = 0.86), and 99 students were at Time 3 ($M$ = 2.02, $SD$ = 1.14). The means of both classes increased over the three time points, but the greatest increase for both classes was from Time 2 (Pretest 2) to Time 3 (posttest). The effect of the EAI instruction was clearly stronger in Class 2, the higher ability class, than in Class 1. In fact, all but 10 of the students were in Class 2 following EAI instruction.

A repeated-measures analysis of variance was used to analyze possible quantitative differences for patterns 112, 122, 212, and 222. Four of the eight transition patterns—211, 221, 111, and 121—were excluded from the analysis because of very small numbers. An interaction was observed between time point and latent class for transition patterns, $F(6, 190) = 5.65$, $p < .01$, $\hat{\eta}^2 = .151$, which indicates ability across time points changed by transition patterns. The interaction effect was investigated for each time point and each transition pattern separately. For each of the four patterns, a significant change in abilities ($p$s $< .01$, $\hat{\eta}^2 = .751-.996$) was observed across the three time points.

## Discussion

The purpose of this article was to demonstrate the use of the LTA-MRM for assessing possible cognitive changes of students' problem-solving performances after they were taught with EAI. Unlike the usual score-level methods for measuring growth and change, the LTA-MRM provides a way of assessing the effects of an intervention by classifying students into latent classes and detecting patterns of transition between those classes. Members within latent classes are qualitatively homogeneous on the dimension or dimensions along which the latent classes are formed. In addition, the LTA-MRM tracks the patterns of changes in latent class memberships over the period of the study. Depending on the design of the study, the LTA-MRM can be used to assign the cause(s) of these changes to effects of the intervention.

In this article, we applied the LTA-MRM to a reanalysis of data from a previous study (Bottge et al., 2007) for the purpose of detecting possible performance differences in response to EAI of students who were considered high achieving (pre-algebra), typically achieving, and low achieving (i.e., students with LD). The application of the LTA-MRM showed the same overall effects as those found in the Bottge et al. analysis. For example, both analyses indicated that students with LD scored higher on the posttest than on the pretests and benefited from EAI. However, the application of the LTA-MRM indicated that there were qualitative as well as quantitative differences in student growth as a result of the EAI intervention that were not evident from the original analysis. For example, initial differences in performance indicated that students in Class 2 were better able to answer the complex mathematics items than students in Class 1. Transition patterns, which showed the effects of the EAI intervention, also indicated that students in both latent classes responded positively to EAI. This was particularly evident in the case of students who were initially in Class 1 but who changed to Class 2 and were much more successful in solving the more complex mathematics items following EAI.

Perhaps most important was the impact of EAI on the problem-solving performances of the 17 lowest achieving students and the 8 students with LD. The LTA-MRM analysis showed that these 25 students moved from the low-ability class (i.e., Class 1) to the high-ability class (i.e., Class 2) following the EAI intervention. Students who scored at the lower ability levels just prior to the intervention had a very high probability of .826 of moving into the group of high-ability students following EAI. This is a particularly encouraging finding because the FOC test required students to navigate problems and accomplish several complex tasks such as interpreting a schematic plan, figuring out the most economical combinations of lengths of wood, and computing the total cost of the project.

The LTA-MRM also aided in identifying strengths and weaknesses of students with LD on particular items. For example, the total test score of one student with LD across time points did not seem to indicate any improved learning from EAI. However, the new analysis revealed that he answered several of the more complex items correctly and moved from Class 1 to Class 2 after the EAI intervention. In contrast, the overall score of another student with LD was quite high, but the LTA-MRM showed he could not answer some of the more complex items. Identifying these cognitive changes would be a special help to teachers as they plan remediation for individual students.

### Limitations

The results of this study show promise for developing new ways of interpreting what students learn from instruction, but limitations should be noted. First, the host study used the same set of performance tasks over all three time points. Although students were not told their results, it is possible that memory may have played a part in their responses to test items. Memory effects, response consistency effects, and practice effects are all potential problems that may exist in designs with repeated measures. In this study, we found evidence of memory effects in one of the transition patterns. Thus, it is possible that the local independence assumption could have been violated, in part, within some response patterns.

It is also possible that ability estimates reflected a somewhat different response process in the two latent classes because items requiring two or more skills were clearly more difficult for members of Class 1 than for Class 2. Although these items function differently across classes, the presence of differential item functioning does not necessarily imply that the constructs being measured in each latent class are different (Maij-de Meij, Kelderman, & van der Flier, 2008). However, as Embretson and Reise (2000) note, it is still not clear whether abilities should be interpreted the same across classes or whether it is necessary to adjust scores for deficient skill states.

### Implications for Practice

Despite calls for research on assessments that can improve learning, we currently have available few, if any, strategies for helping low-achieving adolescents develop the cognitive representations needed to make sense of the types of mathematics items typically included in high-stakes assessments (e.g., Bransford, Brown, Cocking, Donovan, & Pellegrino, 2000). Typical test accommodations for students with disabilities include extended time and oral reading of test directions or items, but neither strategy significantly helps students with math disabilities understand the concepts addressed in test questions (Sireci, Scarpati, & Li, 2005). Although accommodations discussed in the *Standards for Educational and Psychological Testing* (American Educational Research Association, American Psychological Association, & National Council on Measurement in Education, 1999) are intended to increase the validity of test results for students with disabilities, interpretation of scores on tests administered under inconsistent conditions is problematic.

In this study we showed how LTA-MRM can be used to examine the effects of instruction that go beyond the score level and manifest variables, such as membership in a particular demographic group or a function of ability or achievement level. Mixture IRT models can detect latent groups based on homogeneities in the patterns of responses to test items used by members of a latent group, as they did in this study. In this regard, substantial information was obtained by a careful analysis of patterns of responses by latent groups of individuals to test items (Mislevy & Verhelst, 1990). Previous research has demonstrated that latent groups of students who differ can be identified in the kinds of cognitive strategies they use to answer test questions (Bolt, Cohen, & Wollack, 2001; Embretson & Reise, 2000; Rost, 1990, 1997).

Analysis methods such as LTA-MRM are important for building new teaching and measurement models. First, the new model can guide development of appropriate instructional tools and evaluate how they are working, especially among low-achieving populations such as students with LD. This would benefit teachers and administrators who are looking for ways to assess students' RTI. Second, the model can incorporate new ways of helping teachers consider both the potential heterogeneity in RTI and the effects of an intervention over time. This would make test results more meaningful to teachers and address some of the concerns described in the Mathematics Advisory Panel report. Finally, the model can give more credence to researcher-developed tests and provide a more sensitive alternative to standardized tests, thereby helping to eliminate false negatives in developmental research.

## Declaration of Conflicting Interests

The author(s) declared no potential conflicts of interest with respect to the authorship and/or publication of this article.

## Funding

The author(s) disclosed receipt of the following financial support for the research and/or authorship of this article: This work was supported in part by the U.S. Department of Education, Office of Special Education and Rehabilitative Services, Technology and Media Services for Individuals with Disabilities (H327A020030-04) and the U.S. Department of Education, Institute of Education Sciences Cognition and Student Learning Research Program (R305H040032). Any opinions, findings, or conclusions are those of the authors and do not necessarily reflect the views of the supporting organizations.

## References

American Educational Research Association, American Psychological Association, & National Council on Measurement in Education. (1999). *Standards for educational and psychological*. Washington, DC: Authors.

Andersen, E. B. (1985). Estimating latent correlations between repeated testings. *Psychometrika* , *50*, 3–16.

Barrows, H. L. (1996). Problem-based learning in medicine and beyond: A brief overview. *New Directions for Teaching and Learning*, *68*, 3–12.

Bolt, D. M., Cohen, A. S., & Wollack, J. A. (2001). A mixture item response for multiple-choice data. *Journal of Educational and Behavioral Statistics*, *26*, 381–409.

Bottge, B. A., Heinrichs, M., Mehta, Z., & Hung, Y. (2002). Weighing the benefits of anchored math instruction for students with disabilities in general education classes. *Journal of Special Education*, *35*, 186–200.

Bottge, B. A., Heinrichs, M., Mehta, Z. D., Rueda, E., Hung, Y., & Danneker, J. (2004). Teaching mathematical problem solving to middle school students in math, technology education, and special education classrooms. *Research in Middle Level Education Online*, *27*(1). Retrieved from http://www.nmsa.org/Publications/RMLEOnline/Articles/Vol27No1Article1/tabid/529/Default.aspx.

Bottge, B. A., Rueda, E., Serlin, R. C., Hung, Y. H., & Kwon, J. M. (2007). Shrinking achievement differences with anchored math problems: Challenges and possibilities. *Journal of Special Education*, *41*, 31–49.

Bransford, J. D., Brown, A. L., Cocking, R. R., Donovan, M. S., & Pellegrino, J. W. (2000). *How people learn: Brain, mind and school*. Washington, DC: National Academies Press.

Brown, J. S., Collins, A., & Duguid, P. (1989). Situated cognition and the culture of learning. *Educational Researcher*, *17*(1), 32–41.

Cho, S. J., Cohen, A. S., Kim, S. H., & Bottge, B. (2010). Latent transition analysis with a mixture item response theory measurement model. *Applied Psychological Measurement, 34*, 483-504.

Cognition and Technology Group at Vanderbilt. (1997). *The Jasper Project: Lessons in curriculum, instruction, assessment, and professional development*. Mahwah, NJ: Lawrence Erlbaum.

Cognition and Technology Group at Vanderbilt University. (1990). Anchored instruction and its relationship to situated cognition. *Educational Researcher*, *19*(3), 2–10.

Collins, L. M., & Wugalter, S. E. (1992). Latent class models for stage-sequential dynamic latent variables. *Multivariate Behavioral Research*, *27*, 131–157.

Dale Seymour Publications. (2004). *Connected mathematics*. Upper Saddle River, NJ: Prentice Hall.

Embretson, S. E. (1991). A multidimensional latent trait model for measuring learning and change. *Psychometrika*, *56*, 495–515.

Embretson, S. E., & Reise, S. P. (2000). *Item response theory for psychologists*. Mahwah, NJ: Lawrence Erlbaum.

Fischer, G. H. (1976). Some probabilistic models for measuring change. In D. N. M. de Gruijter & L. J. Th. van der Kamp (Eds.), *Advances in psychological and educational measurement* (pp. 97–110). New York: John Wiley.

Fischer, G. H. (1989). An IRT-based model for dichotomous longitudinal data. *Psychometrika*, *54*, 599–624.

Fischer, G. H. (1995). Linear logistic models for change. In G. H. Fischer & I. W. Molenaar (Eds.), *Rasch models: Foundations, recent developments, and applications* (pp. 157–180). New York: Springer-Verlag.

Fuchs, L. S. (2003). Assessing intervention responsiveness: Conceptual and technical issues. *Learning Disabilities Research & Practice*, *18*, 172–186.

Fuchs, D., & Deshler, D. (2007). What we need to know about responsiveness to intervention (and shouldn't be afraid to ask). *Learning Disabilities Research & Practice*, *22*, 129–136.

Fuchs, D., & Fuchs, L. S. (2006). Introduction to response to intervention: What, why, and how valid is it? *Reading Research Quarterly*, *41*, 93–99.

Gijbels, D., Dochy, F., Van den Bossche, P., & Segers, M. (2005). Effects of problem-based learning: A meta-analysis from the angle of assessment. *Review of Educational Research*, *75*, 27–61.

Greeno, J. G., & the Middle School Mathematics Through Applications Project Group. (1998). The situativity of knowing, learning, and research. *American Psychologist*, *53*, 5–26.

Hoover, H. D., Dunbar, S. N., & Frisbie, D. A. (2001). *The Iowa Tests of Basic Skills (ITBS, Form A)*. Itasca, IL: Riverside Publishing.

Li, F., Cohen, A. S., Kim, S.-H., & Cho, S.-J. (2009). Model selection methods for dichotomous mixture IRT models. *Applied Psychological Measurement*, *33*, 53–373.

Maij-de Meij, A. M., Kelderman, H., & van der Flier, H. (2008). Fitting a mixture item response theory model to personality questionnaire data: Characterizing latent classes and investigating possibilities for improving prediction. *Applied Psychological Measurement*, *32*, 611–631.

Mislevy, R. J., & Verhelst, N. (1990). Modeling item responses when different subjects employ different solution strategies. *Psychometrika*, *55*, 195–215.

Muthén, L. K., & Muthèn, B. O. (2006). Mplus [Computer software]. Los Angeles, CA: Authors.

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

National Mathematics Advisory Panel. (2008). *Foundations for success: The final report of the National Mathematics Advisory Panel*. Washington, DC: U.S. Department of Education.

Pellegrino, J., Chudowsky, N., & Glaser, R. (Eds.), for the National Research Council Committee on the Foundations of Assessment. (2001). *Knowing what students know: The science and design of educational assessment*. Washington, DC: National Academies Press.

Rost, J. (1990). Rasch models in latent classes: An integration of two approaches to item analysis. *Applied Psychological Measurement*, *14*, 271–282.

Rost, J. (1997). Logistic mixture models. In W. J. van der Linden & R. K. Hambleton (Eds.), *Handbook of modern item response theory* (pp. 449–463). New York: Springer.

Sireci, S. G., Scarpati, S. E., & Li, S. (2005). Test accommodations for students with disabilities: An analysis of the interaction hypothesis. *Review of Educational Research*, *75*, 457–490.

Sulzer-Azaroff, B., & Mayer, G. R. (1977). *Applying behavior analysis procedures with children and youth*. New York: Holt, Rinehart & Winston.

von Davier, M. (2001). WINMIRA [Computer software]. St. Paul, MN: Assessment Systems Corporation.

Willett, J. B., & Sayer, A. G. (1994). Using covariance structure analysis to detect correlates and predictors of individual change over time. *Psychological Bulletin*, *116*, 363–381.

Wisconsin Department of Public Instruction. (2006). *Specific learning disabilities eligibility criteria resources*. Retrieved October 23, 2006, from http://dpi.wi.gov/sped/eligild.html

## About the Authors

**Sun-Joo Cho**, PhD, is an assistant professor of Psychology and Human Development at Vanderbilt University. Her current interests include item response theory modeling and its parameter estimation.

**Brian A. Bottge**, EdD, is the William T. Bryan Endowed Chair in the Department of Special Education and Rehabilitation Counseling at the University of Kentucky. His primary interest is developing and testing instructional methods for improving the math performance of low-performing students.

**Allan S. Cohen**, PhD, is a professor of Educational Psychology at the University of Georgia. His interests include psychometric theory and quantitative methodology.

**Seock-Ho Kim**, PhD, is a professor of Educational Psychology and Instructional Technology at the University of Georgia. His current interests include item response theory, educational measurement, and applied statistics.