Practical Assessment, Research & Evaluation

A peer-reviewed electronic journal.

Volume 14, Number 16, October 2009 ISSN 1531-7714

Diagnostic Assessments in Mathematics to Support Instructional Decision Making

Leanne R. Ketterlin-Geller
Southern Methodist University
Paul Yovanoff
University of Oregon

Diagnosis is an integral part of instructional decision-making. As the bridge between identification of students who may be at-risk for failure and delivery of carefully designed supplemental interventions, diagnosis provides valuable information about students' persistent misconceptions in the targeted domain. In this paper, we discuss current approaches to diagnosis in mathematics and highlight the strengths and limitations of each approach for making instructional decisions. We point to cognitive diagnostic assessments as an emerging solution for providing detailed and precise information about students' thinking that is needed to provide appropriate educational opportunities for students struggling in mathematics.

In this paper, we focus on defining current approaches to diagnostic assessment in mathematics and discuss the utility of the results for guiding instructional design and delivery decisions for students at-risk for failure. The purpose of this article is to help practitioners determine the most appropriate type of diagnostic approach given the intended decisions. To this end, we define the current conceptualizations of diagnosis in practice and discuss their relative strengths and limitations. We highlight the value of cognitive diagnostic assessment for making instructional decisions and describe why, in these authors' opinion, this approach may be the best available method for supporting student achievement through the design of supplemental interventions for struggling students. We illustrate fundamental issues with a series of examples using multiplication and division of fractions.

Differing Definitions of Diagnosis

In education, diagnosis assumes different meanings and is frequently approached from different perspectives. Considerable variability exists with respect to the definition of diagnosis in education. From a clinical perspective, diagnosis may assume a medical definition in which assessment results are used to determine the likelihood of a specific condition. For example, in special education, a school psychologist or other licensed and qualified practitioner evaluates standardized educational and psychological assessments to classify a student as having a learning disability. Until the recent reauthorization of the Individuals with Disabilities Education Act (IDEA; 2004), the most frequently used criteria for this diagnosis was the discrepancy between achievement and results on a standardized measure of intelligence.

Alternatively, diagnosis may assume an instructional definition in which assessment results provide information about students' mastery of relevant prior knowledge and skills within the domain as well as preconceptions or misconceptions about the material. Teachers use this information to adjust instruction by identifying which areas students have and have not mastered. This results in varied instructional plans that are responsive to students' needs (Fuchs, Fuchs, Hosp, & Hamlett, 2003). However, the time involved in administering, interpreting, and implementing changes based on these approaches may cause many educators to avoid using diagnostic tests to guide instructional decisions (Oosterhof, 2003).

In addition to the perceived lack of efficiency of diagnostic assessment, there is general confusion over the types of assessments that can be used for diagnosis. In K-12 mathematics, two types of assessment practices are currently used to provide diagnostic information: response analyses and cognitive diagnostic assessments. Response analysis is based on students' responses to instructionally-relevant item sets and provides ongoing information about students' mastery and/or application of current knowledge and skills. Analyzing students' responses to problems can be used to adjust instruction so as to correct students' *current* misunderstandings; however, limited information about students' persistent and systematic thinking errors may be tendered from these analyses. Conversely, cognitive diagnostic assessments have the potential to provide appraisals of specific student-level cognitive processes that are structured on the basis of cognitive theory and statistical modeling of response patterns. This information can be used to provide valuable instructional information needed to design remedial instructional programs or supplemental interventions. To help practitioners differentiate between these assessment techniques and select the most appropriate tool for their uses, we describe each approach and discuss their relative strengths and limitations for making instructional decisions (see Table 1 for a summary). We highlight the value of cognitive diagnostic assessments for designing supplemental instructional interventions for students who are struggling.

Table 1. Comparison of diagnostic assessment approaches.

Diagnostic approach	Instructional use	Content referent	Score estimation	Classification
Cognitive Diagnostic Assessment	Identify persistent misconceptions to design supplemental instruction/ interventions	Theory of cognitive processing in domain	Knowledge state	Mastery of multidimensional cognitive attributes
Skills Analysis	Identify skills that may be problematic to design review activities	Broad skills across the curriculum	Skill aggregation	Mastery of unidimensional subskills
Error Analysis	Identify errors students are making when solving specific problem types to design reteaching sequences	Procedural knowledge across the curriculum	Distractor analysis	Error patterns

ROLE OF DIAGNOSTIC ASSESSMENTS IN INSTRUCTIONAL DECISION MAKING

Clarifying the definition of diagnosis and diagnostic assessment is underscored by the critical role diagnosis plays within an instructional decision-making model. In an integrated assessment-instruction system, all students are screened approximately three times per year to determine which students are on-track for success and which students may be at-risk for failure in the domain (Ketterlin-Geller, Baker, & Chard, 2008). Once students are classified by risk status, students who have a high probability of not meeting the outcome goal are administered diagnostic assessments. It is assumed that students identified in this category have persistent deficits in their knowledge or skills that preclude successful engagement in the core curriculum. As such, students at-risk for failure typically receive supplemental instructional interventions designed to overcome these deficits. To determine the domain-specific topics in which remediation is needed, diagnostic assessments are administered to these students (Stecker & Fuchs, 2000). To aid in instructional design, diagnostic tests should measure students' competencies on components embedded within the theoretical model of learning (Gregoire, 1997). Such diagnostic assessments identify specific deficits or persistent misconceptions in students' requisite pre-skills or knowledge. Pre-skills or knowledge include those concepts or tasks that are required in order to successfully complete the targeted tasks within the instructional domain and are often referred to as attributes within the cognitive model (Tatsuoka & Tatsuoka, 1997).

Several assessment models that propose diagnostic inferences are used in the area of K-12 mathematics. Although some of these approaches have been widely used, their utility and psychometric integrity for providing diagnostic information to guide the design of remedial interventions may be limited. These assessment practices typically involve analyzing students' responses through skills analysis or error analysis.

Response Analyses

Typically, response analysis involves teachers', and in some cases students' detailed evaluation of students' answers beyond simple dichotomous scoring of correct/incorrect. Two response analysis techniques are described below: skills analysis and error analysis. These methods differ in their focus and intended use. Skills analysis focuses on strengths and results in an evaluation of students' level of mastery of specific subskills. Error analysis focuses on weaknesses and helps teachers classify students' mistakes. In both cases, assessments elicit responses to specific types of items designed to assist in diagnostic classification. Because of the flexibility in assessment design, these diagnostic procedures can be applied to a variety of tasks including homework, classroom-based quizzes, or standardized tests.

Skills Analysis. Skills analysis involves aggregating student's item-level responses to determine skill mastery associated with specific subskills. In mathematics, skills analysis is emerging as a means for diagnostic interpretation of curriculum-based measures (CBM) (Fuchs & Fuchs, 1990). CBM has a long history as a technically adequate measurement tool for students with special needs (Lembke & Stecker, 2007). CBM is an efficient system for gathering reliable information about student performance using quick probes that are easy to administer and score. As a measurement system, CBMs have been widely used in the areas of reading, spelling, writing, and mathematics as screening tools to identify students who may be at risk for failure in the domain. Additionally, CBMs have been used as progress monitoring tools for evaluating students' rate of growth. Over the past three decades numerous research studies have substantiated the appropriateness of these uses of CBM results (Fuchs, 2004).

Because of the ease of use and efficiency of mathematics CBMs, researchers have recently begun to explore the diagnostic capabilities of these measures by conducting skills analyses from student performance data. Skills analysis refers to the aggregation of performance data for different subskills in order to create students' skills profiles (Fuchs & Fuchs, 1990). Skill profiles describe students' mastery of the knowledge and skills in the tested domain. Although some studies indicate increased student achievement and better delineated instructional plans when teachers use skills analyses (Fuchs & Fuchs, 1990; Fuchs et al., 1994), several constraints in the assessment model may prohibit accurate cognitive diagnosis of student pre-skills and knowledge.

From a psychometric perspective, CBMs have limited utility for making diagnostic decisions because of the domain sampling techniques used to create the measures. CBMs are most commonly created by sampling skills and knowledge representative of the year's curriculum (Lembke & Stecker, 2007). Subsequent alternate forms mirror these specifications. Although this procedure may be appropriate for making screening and progress monitoring decisions, in the authors' opinion, several problems arise from this sampling approach when trying to make diagnostic inferences from subscores. First, because the year's curriculum is broadly sampled to create the test blueprint, essential knowledge and skills in the targeted construct may be under-represented. Construct under-representation occurs when the sampling plan insufficiently represents or reduces the content or cognitive complexity of the targeted construct (Downing & Haladyna, 2004). When behavior is sampled with only a few items per sub-skill per CBM probe, the target skills are likely under-represented. Furthermore, adequate sampling of student behaviors is compromised by CBM administration procedures. CBMs in mathematics are typically administered under timed conditions ranging from 1-6 minutes. Within this time span, most students (by design) are not able to respond to all items, thereby further limiting the sampling of student ability across the subskills or knowledge and limiting the diagnostic inferences made from subscore analysis.

An additional concern when making diagnostic decisions based on skill analysis of mathematics CBM results that arises from this sampling approach is subscore unreliability. As noted by Christ, Scullin, Tolbize, and Jiban (2008) "variability in test material decreases the dependability of measurement outcomes, because the number of items that represent specific domains is uncontrolled and inconsistent" (p. 203). Investigating this issue using simulated data, Miller (2008) projected subscore reliabilities for assessment systems with varying overall reliability coefficients. For a hypothetical test that has an overall reliability of $r = .85$, common of many mathematics CBMs, the subscore reliability with five subscores drops to $r = .53$ and with 15 subscores the reliability is $r = .27$. For diagnostic purposes, multiple subscores may be needed to design appropriate instructional programs to remediate specific knowledge and/or skill deficits. Although the criteria for the stability of test scores used for low stakes decisions such as diagnosis is considerably more flexible than for higher-stakes decisions (Harlen, 2007), Miller's research highlights the impact on score reliability when reporting subscores at the level necessary for diagnostic decisions. As documented by Lyrén (2009), the utility of subscores can be evaluated using other methods such as classical test theory analyses of the reliability of the observed subscores as predictors of true subscore and total performance. To support the use of skills analysis using CBM, the value added information of subscores for making diagnostic decisions should be empirically evaluated in operational administrations.

Consider a typical mathematics CBM for grade 7 students illustrated in Figure 1. Skills measured on this probe represent the year's curriculum and include addition, subtraction, multiplication, and division of complex whole numbers and rational numbers. Students have an average of 4 minutes to complete as many items as possible. The score report for this probe indicates the number of items on the probe that measure each subskill, the student's score and associated skill analysis. As indicated, the summary judgment for the skill analysis is based on minimal student response data. Although skills analyses may provide teachers with useful information for structuring judicious review to support subskill mastery, student performance on this probe may or may not be indicative of persistent misconceptions that need remedial instruction. Moreover, no information is provided about what components of the subskills are problematic or why the student missed the problem. As such, the utility of skills analysis for designing supplemental instruction may be limited.

SAT® Skills Insight™ is another example of the skills analysis approach to diagnosis. Designed as a self-assessment system, SAT® Skills Insight™ elicits students' perceived mastery of college-preparatory knowledge and skills. For each content area domain, students review qualitative descriptors and sample items targeting the knowledge and skills associated with specific score bands. Students evaluate their proficiency on these academic skills by assessing the relative ease of the sample items. Suggestions for skill improvement are provided. As with teacher-driven skills analysis techniques, results from the SAT® Skills Insight™ can help guide the selection of content for review to support subskill mastery. However, detailed diagnostic information about students' underlying misconceptions in the subskills is not provided.

Error analysis. Another commonly used method for identifying students' misunderstanding in mathematics is error analysis. Error analysis is the process of reviewing student's item responses to identify a pattern of misunderstanding. Errors can be classified into two categories: slips and bugs. Slips are random errors in students' declarative or procedural knowledge that are not the result of inherent misunderstandings in the domain. Bugs represent persistent misconceptions about domain specific knowledge or skills that consistently interfere with students' demonstration of their abilities. Identifying bugs, i.e., persistent errors in student thinking, is the primary interest of diagnostic assessment.

As an example of error analysis in mathematics, Ashlock (1994) classified computational-skill bugs into three basic categories: (a) wrong operation, in which the student uses an inappropriate operation when attempting to solve a math problem, (b) computational or fact error, in which the student uses the appropriate operation but makes an error involving basic number facts, and (c) defective algorithm, in which the student uses the appropriate operation but makes a non-number fact error in one or more steps of applying the strategy or selects an incorrect strategy. As an example of the defective algorithm error for a division of fraction problem, a student might correctly invert the divisor but then cross-multiply as though the problem were equated. This error represents a misunderstanding when

Figure 1. Typical Mathematics Curriculum Based Measure (CBM).

$6\overline{)39}$	Convert to decimal: $\frac{1}{3} =$	$\frac{8}{9}+\frac{4}{3}$	$\frac{1}{4}\div\frac{2}{5}$	$29.41 - 8.67$
$\frac{2}{3}\div\frac{8}{9}$	$19.04 - 4.02$	450×29	Convert to fraction: .7 =	$\frac{3}{4}\times\frac{5}{9}$
$\frac{7}{5}+\frac{3}{5}$	$5\overline{)672}$	Convert to fraction: .4 =	398×31	$381.2 + 784.1$
$\frac{4}{7}-\frac{2}{5}$	38.1×8	$\frac{1}{9}\times\frac{7}{8}$	$65.25 - 9.37$	Convert to decimal: $\frac{1}{5} =$
Convert to decimal: $\frac{1}{4} =$	$2\overline{)9.2}$	$\frac{6}{11}\div\frac{1}{2}$	$\frac{2}{5}+\frac{1}{3}$	Convert to fraction: .5 =

Score Report for Zachary

Skill	Number of Items Per Skill	Zachary's Score	Zachary's Skill Analysis
Long division	2	2	●
Convert to decimal	3	1	⭘
Convert to fraction	3	2	⭘
Multiplication with carrying	2	1	⭘
Addition of fractions	3	1	⭘
Subtraction of fractions	1	0	○
Multiplication of fractions	2	1	⭘
Division of fractions	3	1	⭘
Addition with decimals	1	1	●
Subtraction with decimals	3	1	⭘
Multiplication with decimals	1	1	●
Division with decimals	1	0	○

● = Mastered
⭘ = Partial Mastery
○ = Not Mastered

applying one component of the "invert and multiply" strategy for solving division of fractions problems. Additional errors associated with solving story-based problems include those that involve interpreting and applying the language, such as decoding, vocabulary, and translation of the text to number sentences.

Figure 2 presents possible student responses to a sampling of items from the CBM probe presented in Figure 1. These responses can be examined and classified by error type (see Figure 2). Although error analysis can provide timely information for adjusting instruction so as to avoid reinforcing incorrect procedures, this information may not provide insights into the cognitive attributes students have or have not mastered that form the basis for designing remedial instruction or supplemental interventions. Instead, teachers often focus on correcting the procedural errors that are evident from error analysis without recognizing the conceptual understanding that provides the foundation for skill application (Russell & Masters, 2009). Additionally, teachers may need to aggregate large samples of student performance data to determine if the error is a random slip or a persistent bug. This classification has considerable implications for instruction and may determine if the student will benefit from reteaching or needs remedial instruction.

Although skills and error analyses may provide useful information about students' responses to the current classroom instructional sequence, these response analysis techniques have limited utility for making decisions about students' underlying cognitive processing. To arrive at a diagnostic decision about subskill mastery, these response analytic techniques assume that subskills are unique and independent. As

Figure 2. Error Analysis for CBM Probe.

Item	Possible Student Response	Error Analysis
$6\overline{)39}$	$\begin{array}{r} 6.3 \\ 6\overline{)39} \\ \underline{-36} \\ 3 \end{array}$	Defective algorithm (component error)
$\frac{8}{9}+\frac{4}{3}$	$\frac{8}{9}+\frac{4^{\times 3}}{3_{\times 3}} = \frac{8}{9}+\frac{12}{9} = \frac{20^{\div 2}}{18_{\div 2}} = \frac{10}{9}$	Defective algorithm (component error)
$\begin{array}{r} 450 \\ \underline{\times\ 29} \end{array}$	$\begin{array}{r} {}^{4}450 \\ \underline{\times\ 29} \\ 4050 \\ \underline{800} \\ 12050 \end{array}$	Computation or Fact error
$\frac{3}{4}\times\frac{5}{9}$	$\frac{3}{4}\times\frac{5}{9} = \frac{3}{4}\times\frac{5}{9} = \frac{27}{20}$	Wrong operation
Convert to fraction: .4 =	Convert to fraction: $.4 = \frac{40}{10}$	Defective algorithm (strategy error)
$\begin{array}{r} 381.2 \\ \underline{+\ 784.1} \end{array}$	$\begin{array}{r} {}^{1}381.2 \\ \underline{+\ 784.1} \\ 2265.3 \end{array}$	Computation or Fact error

such, measurement of these skills can occur in isolation of other skills. This unidimensional approach makes classification of student mastery relatively straightforward. Many skills, however, do not develop in isolation of others. As proposed by some learning theorists (cf., Bransford, Brown, & Cocking, 2000), cognitive processes leading to domain mastery may be dependent upon concurrent development of multiple skills or attributes. It follows that an item response may be the result of various combinations of skill strengths and weaknesses. This multidimensional network of skills underlying the cognitive model makes the unidimensional process of response analysis impossible. In these instances, more complex modeling of student responses is needed to provide diagnostic information for intervention design.

COGNITIVE DIAGNOSTIC ASSESSMENTS

An emerging approach to diagnosis for instructional decision-making relies on cognitive models of learning to determine students' persistent cognitive errors. Because cognitive models are based on empirical research on learning, they provide a foundation for understanding the pre-skills and knowledge involved in successfully engaging with the material (Pellegrino, Chudowsky, & Glaser, 2001). This foundation is used to structure remedial instructional opportunities and supplemental interventions for students with specific cognitive errors.

As an introduction to the need and design of cognitive diagnostic assessment for instructional design, it is worthwhile to note briefly some historical developments. Cognitive diagnosis is the merger of two major research fields, (a) cognitive psychology, and (b) psychometric modeling. The resulting field of cognitive diagnostic measurement is a relatively current development.

Role of Cognitive Psychology

Cognitive diagnosis requires the identification of the cognitive attributes that can be combined to form knowledge states underlying observed performance. Cognitive attributes are domain-specific pre-skills and knowledge that are needed to demonstrate mastery in the targeted construct (Chipman, Nichols, & Brennan, 1995; Leighton & Gierl, 2007). The cognitive model is a differentiating feature of this approach, and can be seen as "an architecture organizing the successive processes involved" in learning (Gregoire, 1997, p. 17). Attributes are typically isolated through careful task analyses, expert review, verbal protocols, and other inquiry methods for analyzing student thinking processes (Gorin, 2007). Once the attribute structure for the cognitive model has been determined, combinations of attributes that make up students' knowledge states can be identified.

Knowledge states are well-specified combinations of attributes that form the basis of students' conceptions of domain-specific knowledge and skills. Knowledge states represent the level of mastery of a unique combination of attributes that characterize specific misconceptions or cognitive errors, ranging from competence in none to all of the attributes within the cognitive model. Theoretically, it is possible to have a large number of knowledge states depending on the number of attributes that can be combined. In practice, however, because students often approach problem solving in the domain with similar misconceptions, there are a finite number of plausible and testable combinations. Furthermore, the cognitive model constrains the class of theoretically reasonable knowledge states.

Because knowledge states underly students' persistent (mis)conceptions within the cognitive model, these form the basis for designing supplemental instructional modules for remediating these deficits. Without this precise intersection between cognitive diagnosis and instructional design, it is the authors' opinion that students at-risk for failure in the domain may not receive the necessary instructional supports needed to remediate their deficits or misconceptions. As such, cognitive diagnostic assessments are needed to maximize the learning potential for all students.

Role of Psychometric Modeling

Dominant psychometric models developed over the past 50 years in educational measurement tend to provide elegant solutions for item/test development, item parameter calibration, and accurate examinee scaling on unidimensional and multidimensional traits that are useful for developing cognitive diagnostic assessments. For instance, item response theory (IRT) and latent class modeling have resulted in an explosive amount of research over the past 50 years. With the advent of new estimation algorithms and desktop computing power, new and highly flexible psychometric models relating test responses to latent trait scales are routinely proposed in measurement journals. For example, Rudner and Talento-Miller (2007) applied Bayes' theorem of inverse probabilities (Press, 1989) to make diagnostic inferences based on response analysis procedures. Using items with known item response theory (IRT) psychometrics (e.g., item difficulty, item discrimination, item guessing), the Bayesian procedure requires a priori estimates of probabilities that a randomly sampled student will be in any one of the diagnostic classification categories. Also, the procedure requires a priori estimates of item response probabilities given a mastery category. Posterior mastery classifications are made based on the (a) the observed scored response pattern, and (b) estimated priors (probability distribution of classification categories, probability of item response given mastery classification). As noted earlier, the response analysis application of diagnosis assumes a unidimensional trait structure in which items are associated with one, and only one, skill. However, when the purpose of diagnosis is to evaluate students' cognitive processing in domains that represent combinations of skills, more complex item sampling and statistical models are needed to make accurate diagnostic inferences.

A variety cognitive diagnostic measurement models require a cognitive model delineating the cognitive attributes underlying performance in a specific achievement domain (Leighton & Gierl, 2007). Because most attributes cannot be tested in isolation, most items address a combination of attributes. For each item, the tested attributes are recorded in a Q-matrix. A Q-matrix provides an index for cataloging which items measure specific attributes. In a Q-matrix, attributes *k* (rows) are related to items *i* (columns). Referencing the Q-matrix makes it possible to classify a student's knowledge state based on his or her observed item response pattern. Because it is assumed that when a student answers an item or series of items correctly he or she has mastered the attributes associated with those items, cross-referencing the Q-matrix with the student's response pattern provides a map of the student's mastered and non-mastered attributes. This classification can subsequently be used to design remedial instruction or supplemental interventions.

Increasing numbers of creative and flexible cognitive diagnostic models appear in the literature and at national conferences. Generally, the models hypothesize an underlying latent trait and/or latent class structure, and can be differentiated based on their model constraints, assumptions, and most suitable application. Some models require mastery of sets of relevant attributes for successful item response (conjunctive models), while other models are not quite as restrictive (disjunctive models) (For a comprehensive summary of different cognitive diagnostic assessment models, see Rupp and Templin (2007) and Fu and Li (2007)).

The item-attribute representation implied by the cognitive model marks the key distinction between this diagnostic approach and response analysis techniques. In response analyses, subskill scores are obtained by aggregating performance on items that measure only one skill. In contrast, items written for cognitive diagnostic assessments measure an array of interrelated attributes based on the cognitive model, thereby precluding simple aggregation of results to arrive at a diagnostic classification. As such, cognitive diagnostic assessments model a multidimensional problem in which the conjunctive or disjunctive association between attributes influences item performance and subsequent diagnostic classification.

A simplified example case is presented in Figure 3 that depicts a sample of diagnostic items for division of fractions and illustrates the cognitive model embedded within a classification matrix. Student performance is illustrated in the figure. Although logical reasoning can help teachers identify students' misconceptions, measurement modeling is more efficient and precise. By modeling if the observed errors are merely 'slips' or identifiable 'bugs,' the specificity of classification provides teachers with a clear indication of what aspects of the target skill students have or have not mastered. As such, teachers can use this information to design or select supplemental instruction tailored to individual needs.

INTEGRATING MULTIPLE APPROACHES TO DIAGNOSIS

Combining diagnostic assessment approaches may also prove useful for designing instructional programs to remediate students' misconceptions. By integrating the principles of cognitive psychology with response analysis, diagnostic assessments can be created to provide insights into persistent errors that interfere with student learning in the targeted domain. In this approach, multiple choice items can be strategically designed to incorporate distractors that mirror systematic errors in student thinking. Using distractor analysis, students' responses are aggregated to determine persistent misconceptions across items. Other item types can be similarly designed to test specifically for important, instructionally relevant errors.

This approach to diagnosis has shown promising results in mathematics. In a randomized controlled study

Figure 3. Sample cognitive diagnostic items and classification matrix for division of fractions.

	Items																			
Cognitive Attributes	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
Conceptual understanding of	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X
Ability to convert mixed number				X	X	X	X				X			X			X	X		X
Ability to multiply fractions							X	X	X	X	X	X	X	X	X	X	X	X	X	X
Conceptual understanding of the division											X	X	X	X	X	X	X	X	X	X
Ability to apply the invert and															X	X	X	X	X	X
Zachary's Responses (C = correct; I = incorrect)	C	C	C	I	C	C	I	C	C	C	I	C	C	C	C	I	I	I	I	I

Summary classification for Zachary:

Attributes Mastered:

- Conceptual understanding of fractions
- Ability to multiply fractions
- Conceptual understanding of the relationship between multiplication and division

Focus of Supplemental Instruction (attributes not mastered):

- Ability to convert mixed number to improper fraction
- Ability to apply the invert and multiply algorithm

targeting three common misconceptions in algebra, Russell, O'Dwyer, and Miranda (2009) found that students participating in an integrated diagnostic assessment and instructional intervention performed significantly better on a measure of algebra proficiency than did students participating in typical classroom instruction without guidance from diagnostic information. Although overall algebra achievement increased, there was no statistically significant effect on the presence of specific misconceptions due to group membership.

A possible explanation for these findings might relate to the analytic procedures used to determine students' misconceptions. In contrast to the cognitive diagnostic measurement models previously described, this approach to diagnosing students' misconceptions does not account for sampling error in test design. Integrating cognitive diagnostic measurement models that estimate slipping and guessing parameters at either the item or attribute level when making diagnostic classifications may account for variability in students' scores in relation to the persistent misconceptions as opposed to random errors.

CONCLUSIONS: VALIDITY OF DIAGNOSTIC DECISIONS

Within an instructional decision-making model, diagnostic test results are increasingly used to guide the design of remedial instruction and placement in supplemental intervention programs. Because these decisions may significantly impact the educational opportunities available to individual students, validity evidence is needed to substantiate test-score use for these purposes. In this article, we highlighted the emergence and utility of cognitive diagnostic assessments for making instructional programming decisions for students at-risk for failure in the domain. The combination of cognitive psychology and psychometric principles in the design of cognitive diagnostic tests may promote valid diagnostic inferences about students' persistent misunderstandings and cognitive errors. Current and emerging research points to these assessment systems as valuable tools to guide instructional design and delivery decisions.

As described in this paper, other assessment systems enable diagnostic inferences based on response analyses of test results. Specifically, skill and error analyses have been used to make some diagnostic decisions in educational contexts. As noted, skill analysis helps classify students' level of mastery of specific subskills, and can be used to design review activities. Similarly, error analysis provides information about the types of mistakes students make to help teachers identify if algorithms or procedures need to be retaught. However, results from these diagnostic techniques may not provide sufficient information about students' cognitive processing in the domain that is needed to design instructional remediation. By carefully considering the validity evidences for each use of an assessment system, over extension of the utility of assessment systems can be averted, thereby circumventing inappropriate decision-making that may result in inadequate services for individuals.

REFERENCES

Ashlock, R. B. (1994). *Error Patterns in Computation, 6th Edition.* Englewood Cliffs, NJ: Prentice Hall.

Bransford, J.D., Brown, A.L., & Cocking, R.R. (2000). *How people learn: Brain, mind, experience, and school.* Washington, D.C.: National Academy Press.

Chipman, S. F., Nichols, P. D., & Brennan, R. L. (1995). Introduction. In P.D. Nichols, S.F. Chipman, & R.L. Brennan (Eds.). *Cognitively diagnostic assessment* (pp. 1-18). Hillsdale, NJ: Lawrence Erlbaum Associates, Inc.

Christ, T. J., Scullin, S., Tolbize, A., & Jiban, C. L. (2008). Implications of Recent Research: Curriculum-Based Measurement of Math Computation. *Assessment for Effective Intervention, 33*, 198-205.

Downing, S. M., & Haladyna, T. M. (2004). Validity threats: Overcoming interference with proposed interpretations of assessment data. *Medical Education, 38*, 327-333.

Fu, J., & Li, Y. (2007). Cognitively diagnostic psychometric models: An integrative review. Paper presented at the Annual Meeting of the National Council on Measurement in Education, Chicago, IL.

Fuchs, L.S. (2004). The past, present, and future of curriculum-based measurement research. *School Psychology Review, 33*, 188-192.

Fuchs, L.S., & Fuchs, D. (1990). The role of skills analysis in curriculum-based measurement in math. *School Psychology Review, 19.*

Fuchs, L. S., Fuchs, D., Hamlett, C. L., Thompson, A., Roberts, P. H., Kubek, P., & Stecker, P. M. (1994). Technical Features of a Mathematics Concepts and Applications Curriculum-Based Measurement. *Assessment for Effective Intervention, 19*, 23-49.

Fuchs, L. S., Fuchs, D., Hosp, M. K., & Hamlett, C. L. (2003). The potential for diagnostic analysis within curriculum-based measurement. *Assessment for Effective Intervention, 28*(3&4), 13-22.

Gorin, J. S. (2007). Test design with cognition in mind. *Educational Measurement: Issues and Practice, 25*(4), 21-35.

Gregoire, J. (1997). Diagnostic assessment of learning disabilities: From assessment of performance to assessment of competence. *European Journal of Psychological Assessment, 13*(1), 10-20.

Harlen, W. (2007). Criteria for evaluating systems for student assessment. *Studies in Educational Evaluation, 33*, 15-28.

Ketterlin-Geller, L.R., Baker, S., & Chard, D. (2008). Assessment strategies for secondary mathematics. In A. Thomas & J. Grimes (Eds.), *Best Practices in School Psychology V*, pp. 465-475. Bethesda, MD: National Association of School Psychologists.

Leighton, J. P., & Gierl, M. J. (2007). Why cognitive diagnostic assessment? In J. P. Leighton & M. J. Gierl (Eds), *Cognitive Diagnostic Assessment for Education: Theory and Applications* (pp. 3-18). New York: Cambridge University Press.

Lembke, E. & Stecker, P. (2007). *Curriculum-based measurement in mathematics: An evidence-based formative assessment procedure.* Portsmouth, NH: RMC Research Corporation, Center on Instruction.

Lyrén, Per-Erik (2009). Reporting Subscores from College Admission Tests. *Practical Assessment, Research & Evaluation*, 14(4). Available online: http://pareonline.net/getvn.asp?v=14&n=4.

Miller, M. D. (2008). Data for school improvement and educational accountability: Reliability and validity in practice. In K. E. Ryan & L. A. Shepard (Eds), *The Future of Test-Based Educational Accountability*, pp 249-261. New York, NY: Routledge.

Oosterhof, A. (2003). *Developing and Using Classroom Assessments* (3rd edition). Upper Saddle River, NJ: Pearson Education Ltd.

Pellegrino, J.W., Chudowsky, N., & Glaser, R. (2001). *Knowing what students know: The science and design of educational assessment*. Washington, DC: National Academy Press.

Press, J. S. (1989). Bayesian statistics: Principles, models, and applications. New York: John Wiley and Sons.

Rudner, L. M., & Talento-Miller, E. (2007, April). *Diagnostic testing using decision theory*. Paper presented at the annual meeting of the National Council on Measurement in Education, Chicago, IL.

Rupp, A. A., & Templin, J. L. (2007). Unique characteristics of cognitive diagnostic models. Paper presented at the Annual Meeting of the National Council on Measurement in Education, Chicago, IL.

Russell, M., & Masters, J. (2009). *Formative Diagnostic Assessment in Algebra and Geometry*. Paper presented at the Annual Meeting of the American Education Research Association, San Diego, CA.

Russell, M., O'Dwyer, L, & Miranda, H. (2009). Diagnosing Students' Misconceptions in Algebra: Results from an Experimental Pilot Study. *Behavior Research Methods, 41*(2), 414-424.

Stecker, P. M., & Fuchs, L. S. (2000). Effecting superior achievement using curriculum-based measurement: The importance of individual progress monitoring. *Learning Disabilities Research and Practice, 15(3),* 128-134.

Tatsuoka, K. K., & Tatsuoka, M. M. (1997). Computerized cognitive diagnostic adaptive testing: Effect of remedial instruction as empirical validation. *Journal of Educational Measurement, 34*(1), 3-20.

Citation

Ketterlin-Geller, Leanne R. and Yovanoff, Paul (2009). Diagnostic Assessments in Mathematics to Support Instructional Decision Making. *Practical Assessment, Research & Evaluation*, 14(16). Available online: http://pareonline.net/getvn.asp?v=14&n=16.

Corresponding Author

Leanne Ketterlin Geller, Ph.D.
Department of Education Policy and Leadership
Southern Methodist University
PO Box 750381
Dallas, TX 75275-0381
Phone: 214-768-4947
Email: lkgeller [at] smu.edu