Error Identification Assessments: Diagnostic Instruments *for* Instruction

by
Alvin F. Larson, Ph.D.
Maureen E. Kelleher, Ph.D.

Paper presentation at the 2009 AERA Convention
Division H: Research, Evaluation and Assessment in Schools
Section 3: Assessment in the Schools
Symposium: Common and Day-to-Day Classroom Assessment: Fostering Formative Uses by Teachers and Students

Abstract

Current high-stakes national and state multiple-choice assessments have been designed to service administrative decision makers, but these tests do not provide timely or diagnostic information to teachers. In an effort to be instructionally informative, a new generation of assessments is being developed: Cognitive Diagnostic Assessments (CDA). The diagnostics of CDA are partly based upon what researchers have called the identification of each student's pattern of errors. The demand for diagnostic assessment for education has prompted one local school district to develop, with teachers, its own version of a diagnostic assessment system. These Error Identification assessments (EIa) include many of the CDA concepts into an integrated set of district-wide assessments in math and reading comprehension. The EIa scoring and reporting system is designed to promptly identify the type of error each student is making to help inform each teacher's on-going instruction in grades 2 to 9. This includes the development of a new reading comprehension error vocabulary and the use of the EIa items for instruction. Preliminary qualitative results indicate these diagnostic instruments are valued and utilized by teachers and parents.

There is a demand for Cognitive Diagnostic Assessments (CDA) in K-12 education. This demand originates from two sources: assessment developers who are arguing for radical shifts in how assessments are designed (Huff & Goodman 2007) and teachers, who want timely results that are meaningful to them. Teachers need valid diagnostic information *during* instruction and designed *for* their instruction. Gierl and Leighton (2007) write that there are increasing calls "for instructionally relevant information that can be gleaned from what students think about and do when they solve items on tests". However, for diagnostic assessment, a "fine grain" size is required to make inferences about a student's specific cognitive skills which will also require more frequent testing.

The use of diagnostic tests has historically been applied to psychological assessment and diagnosis, with few examples of educational tests designed for this purpose. Recently, educators have recognized educational assessments as missed opportunities to inform educational decisions. This realization is evident in the No Child Left Behind (NCLB) Act of 2001 in the United States, as it pertains to the development and use of yearly standardized achievement tests (Gorin, 2007).

> Such assessments shall produce individual student interpretive, descriptive, and diagnostic reports … that allow parents, teachers, and principals to understand and address the specific academic needs of students, and include information regarding achievement on academic assessments aligned with State academic achievement standards, and that are provided to parents, teachers, and principals as soon as is practicably possible after the assessment is given. (NCLB, Part A, Subpart 1, Sec 2221[b]3[C][xii], 2001).

Using the concepts of CDA to improve instruction has been piloted for years in laboratory and selected math and science classes but there are beginnings in other school subject areas (Snow & Lohman, 1993). At the college level Mestre, Gerace, Dufresue and Leonard (1997) piloted "active-learning strategies" in introductory physics classes, where the focus was on the generation of questions to explore student errors in reasoning and misconceptions. Also, Bennett (1993) has examined studies where wrong answers to complex problems are analyzed into levels of errors. Additional research in mathematics has also indicated that many faulty student cognitive processes go undetected by classroom teachers that could be identified through a student's pattern of errors. These misconceptions are "both resilient, persistent and often remain undetected when teachers do not see the highly regular pattern in students' errors, responding to them more as though they were random miscalculations" (National Research Council, 2004).

In grades 3 to 9, the practice of identifying errors to improve instruction in reading comprehension is sporadic and not well defined. The National Research Council has expressed a need for new reading assessments that can help teachers in diagnosing the nature of the problem for a particular student (National Research Council, 2004). The RAND Reading Study Group also advocated a new assessment system for reading comprehension that is designed to estimate each student's thinking and motivation about the reading task. Further "diagnostic, process assessments could help indicate why [each student's measured] reading comprehension is poor" (Snow 2002). Current thinking about reading comprehension creates a demand for new kinds of assessments that reflect the dynamic developmental nature of comprehension and the interactions between the reader, activity, text and context. The tools available to teachers for assessing how well students understand what they read are wanting, so some districts are making their own formative assessments (Block, Gambrell, Pressley, 2002).

This district has piloted its own version of CDAs since the Fall of 2003. These Error Identification assessments (EIa) were developed in an effort to provide timely meaningful information to teachers that could be used to improve their current instruction. These district-wide assessments are currently administered three times per year, and the items were constructed by explicitly designing the foils (distractors) so that they would be appealing to students who have systematic misconceptions or utilize poor or inefficient strategies or behaviors.

Purpose of Error Identification Assessments (EIa)

The purpose of this locally-developed, multi-year, district-wide project is to:

1. develop and expand these CDA applications beyond the laboratory and selected classes to a district level implementation of a system that integrates both instruction and assessments in math and reading comprehension;

2. identify student errors and patterns of errors across multiple-choice items;

3. provide timely computer generated reports with well defined error descriptions that are *meaningful to teachers*; and

4. provide the next recommended instructional step for each student with additional instructional support materials as needed.

Development and Implementation

In 2004 the local board of education adopted a new set of multi-year objectives. One of these objectives was "to develop a system wide assessment system … that would provide meaningful, timely and diagnostic reports to help guide instruction". This district-wide, multi-year project is a practical application of identifying student errors in both math and reading. Without this support from the board and superintendent, the district-wide EIa project would not have been implemented. Selected teachers and administrators helped develop and revise both the assessments and the reports.

In an effort to be "meaningful" to teachers, the district's assessments are less concerned about traditional "ability estimates" and more concerned about each specific error, pattern of errors or demonstrated misconceptions. Each EIa foil is **designed** to attract the poorer reader or less capable math student (Luecht, 2007; Gorin, 2007; Downing & Haladyna, 2006; Haladyna, 2004; Osterlin, 1998; Mislevy, 1993) by providing attractive foils based upon typical errors that classroom teachers actually observe in their classes. EIa assumes that if a student selects a particular pattern of foils, then the student consciously selected those particular foils *because* of his/her misconception(s) or inefficient strategies (Siegler, 2005). These identified errors, verified by teachers, provide insight to inform future instruction.

Context and EIa Production

The NCLB student demographics of this small urban school district in Connecticut consist of about 9,000 students: 2% Asian, 13% Black, 43% Hispanic, 42% White. For the 2008-2009 school year, 56% are eligible for free or reduced priced meals. Student mobility is about 9% within the district and another 9% moving out or into the district each year.

Both the EIa math and reading assessments are printed in an attractive format for both teachers and students by the district's Office of Research and Evaluation using Microsoft Publisher Software. Students may write or underline in the test booklets but they record their answers on a customized scannable answer document (bubble sheet) that is "pre-slugged" (utilizing Design Expert Software) with each student's name, identification number, school and teacher. Students have enough time to answer, (flexible time limits) and they can take out-of-level tests. The completed answer documents are scanned (utilizing Scan Tools Plus Software) and cleaned ("double bubbles" due to poor erasers are identified using SPSS programming) and hand corrected in order to report *what each student was thinking during testing*. SPSS programming produces a series of reports, listing error and pattern of errors by students within classrooms, as well as Individual Error Reports for each student in both math and reading comprehension. These reports require thousands of lines of SPSS code across 52 syntax files per year. Efforts are made to return error reports to teachers within 3 to 5 days.

EIa assessments are currently administered in September for grades 3 to 9, as well as Mid-Year and in May for grades 2 to 9 (an additional EIa administration is currently being developed for the 2009-2010 school year). The EIa "tests" are comprised of 46 original reading comprehension passages, 23 Editing/Revising passages, and 23 math assessments. There are 454 reading comprehension questions, 230 editing/revising questions and about 1,000 math items. Each reading comprehension, editing and math item has one correct, or "most correct" answer, with four foils, each designed to mimic a typical student misconception.

EIa was constructed and revised utilizing traditional approaches to test development. EIa statistically links to the scale scores of the state mandated math and reading tests (American Educational Research Association, *et al*., 1999; Kolen & Brennan, 2004; Feuer, *et al*., 1999). This calibration also provided an EIa scale score and *NCLB* performance labels (Below Basic, Basic, Proficient, Goal and Advanced) for each student. Teachers are familiar with these performance labels, but the main purpose of EIa is error identification. Since EIa also correlates well with state assessments (see Table I), they are used as independent variables in multiple regression equations to predict future performance on state required tests (R range .85 to .93 in reading). The Title I schools in this district are "Targeted Assisted" schools, so each student's predicted achievement on state required testing mandated under NCLB is helpful in determining which students are in danger of not meeting the state's proficiency standards. It should also be noted that this project was partially funded by this LEA's Title I, Part A funds. Therefore, the EIa items and associated programming are available to other LEA's.

TABLE 1

Alpha Reliability and Concurrent Validity Correlations with State Mandated Tests with EIa Assessments in Math and Reading by Grade Level

	Alpha Reliability						Concurrent Validity			
Grade Level	Beginning of Year September 2009		Mid-Year Dec-Jan 2009		End-Year May 2008		Spring 2008		Mid-Year Dec-Jan 2008	
	Math	Reading	Math	Reading	Math	Reading	Math	Reading	Math	Reading
2			.90	.86	.88	.89				
3	.91	.87	.92	.80	.93	.85	.82	.81	.82	.78
4	.92	.86	.92	.82	.93	.88	.85	.81	.84	.78
5	.93	.84	.92	.84	.93	.86	.88	.80	.87	.80
6	.93	.88	.92	.86	.93	.86	.89	.78	.89	.79
7	.93	.87	.92	.84	.94	.88	.92	.76	.91	.82
8	.93	.84	.91	.85	.91	.84	.89	.86	.90	.80
9	.90	.85	.91	.85	.92	.87				

NOTE: EIa math has 40 to 50 items, reading has 30 items except mid-year grade 2 which has 24 reading items.

Cadre of Reading Teachers

A cadre of teachers received stipends to write original passages and a set of reading comprehension questions. Since teachers do not have a background in assessment, a set of guidelines with specific item and passage protocols with examples was provided to each member of the cadre (about 20 teachers). They were also required to attend an in-service by the Director of the Office of Research and Evaluation. The draft passages and questions were modified and piloted. Unfortunately, the first administration of EIa reading assessments and reports were poorly defined. Teachers and administrators complained about "confusing" results. Assessments were revised again based upon both teacher input and literature reviews. Additional in-services were provided for reading teachers. It would take an additional three years of revision, piloting and informal in-service before original reading passages with acceptable item stems/foils had stabilized with adequate item statistics and concurrent validity with state required reading assessments.

Reading is Different from Math

Identifying the metacognitive errors in reading at the elementary and middle school level is much more complex than identifying elementary and middle school math errors. Math teachers are trained in math algorithms and typical math errors associated with those math processes. If a teacher identifies a student's mistake in math (such as regrouping, common denominator, rounding and place value errors), math teachers are trained in their well established error vocabulary, and they are trained on 'what to do next' (re-teaching that specific math concept). Also, the standardized math assessments are very similar to the text book exercises and lessons. In contrast, this paper proposes that reading teachers in grades 2 through 9:

- do not have an established error vocabulary related to silent reading on standardized assessments; and

- do not have textbook/anthologies that closely match the high-stakes, standardized reading comprehension assessments that utilize multiple choice, inferential questions. These commercially available reading textbook/anthologies are too literal, with the limited multiple-choice inferential questions that are too easy when compared to the state mandated standardized reading assessments which are used to determine each school's NCLB status.

Teacher's Background Knowledge

In order to meaningfully relate assessment results to classroom teachers, a brief background of teacher knowledge about reading is required. In a review of the literature from both the fields of cognitive psychology and reading, metacognition has been defined as the knowledge and control one has over one's learning and thinking, or the ability to monitor one's own learning (Flavell, 1978; Baker & Brown, 1984; Jacobs & Paris, l987). Applied to the reading process, metacognition requires readers to be aware of incoming information to such an extent that they are able to monitor whether or not they understand what is being read, and at the point where comprehension is lost, would have the ability to use self-regulatory (fix-up) strategies to remedy the situation. Proficient comprehension is contingent upon active cognitive engagement which requires that students use metacognition and self-regulatory strategies (Meece, Blumenfeld, & Hoyle, 1988). In order to successfully use self-regulatory strategies, an individual must possess the ability to monitor his or her reading behavior, identify the causes of lost comprehension, and have the skill to use fix-up strategies to remedy the situation. Taken together, metacognitive knowledge and the use of metacognitive strategies facilitate reading comprehension.

When comprehension difficulty occurs, proficient readers are metacognitively aware that something has happened to disrupt their comprehension and they know how to enact a repair strategy such as re-reading or reading ahead (Pearson, Roehler, Dole & Duffy, 1992). Likewise, students' inability to monitor and regulate their reading has a negative effect on their comprehension. Developing an internal monitoring system in which readers can recognize and resolve textural incongruities is essential to proficient comprehension (Baker & Brown, 1984; Pressley, *et al*., 1989). Re-reading, looking back, reading ahead, predicting/revising, and self-questioning can be taught but are only useful when students understand how, when and where to use them. These strategies have also been termed "fix-up strategies" or reading strategies. But regardless of what they are called, or to what extent they have been taught, even with the abundance of research in favor of metacognitive strategy instruction, researchers find little or no strategy use occurring in today's classroom among poor readers (Mayo, l992). Many teachers still struggle with this issue.

Metacognitive knowledge and strategy instruction need to be explicit, intensive and extensive. To achieve this, strategies need to be taught to students directly, over an extended period of time, as part of the existing curriculum. The ultimate goal is to have students using metacognitive strategies automatically, skillfully and appropriately. To do this, they must understand when, where and how to use the strategies they know (Pressley, Woloshyn, *et al*., 1995; Kelleher, 1984). One of the most important metacognitive strategies that has been shown to improve reading comprehension is the "look-back" strategy (Garner, *et al*., 1984). Good readers are more likely than poor readers to "look back" to resolve a problem. However, poor readers can be taught this strategy through explicit instruction, practice and corrective feedback. The general conclusion is that there is a positive relationship between reading comprehension and use of the look-back strategy (Garner, 1987).

We know that good readers are more aware of multiple strategies and are far more likely to change strategies as needed, and to adapt question-answering strategies to the demands imposed by the question (Raphael and Pearson, 1985). Given a choice between answering a question by: (1) going right to the part of the text the question comes from; (2) re-searching the text to find a response that fits the question; or (3) relying on one's prior knowledge, good readers are better able to select the best strategy than poor readers. In addition, good readers are more flexible and adaptable than novices in their use of strategies. They are far more likely to change strategies to meet different reading tasks (Pearson, *et al*., 1992). Children may not know any strategies that yield accurate and fast performance on a given task, but they choose adaptively among the strategies they do know and practice. Students also use multiple strategies at the same time and, with experience, less effective strategies eventually decrease in frequency in favor of more effective strategies (Siegler, 2005). The difficulty that arises is that even if students are taught strategies to remedy their reading comprehension difficulties, teachers cannot determine which students are actually practicing these strategies during silent reading activities, including homework and high-stakes testing. Since the reading teacher cannot *see* or *hear* his/her student's reading and thinking errors during a reading test, the EIa is designed to identify the latent reading errors each student is making during a reading performance. These are the *same* reading errors students are making when reading mathematics, social studies, science or any other reading task. The EIa is primarily designed to promptly report these latent, ineffective reading behaviors so that students receive the proper (explicit, intensive and extensive) metacognitive feedback in a timely manner.

Measuring Reading: the Student's Task

Drawing inferences is an essential part of the on-going comprehension process readers engage in regardless of grade. There is strong evidence that students can learn to improve their abilities to infer as early as second grade (Pearson, *et al*., 1992); However, early reading instruction in Kindergarten through grade 2 primarily depends upon a literal, not inferential, interpretation of the text. Inference is so important that NAEP and current state mandated reading assessments are comprised primarily of inferential questions. EIa, *with* teachers, attempt to identify typical inferential errors and ineffective reading/test taking behaviors. Generally, people believe that a reading comprehension test is about reading a passage and determining the degree students understand that passage. However, how we **measure** the construct of "reading" is dependant, to a large extent, upon the item stems and foils. Reading comprehension is less about the reading comprehension of the passage, and more about the reading comprehension of each item stem, and primarily upon the careful reading comprehension of each **foil** and discrimination between foils**.** The passage becomes simply the context from which to ask questions or perform tasks or exercises.

As students are confronted with challenging reading comprehension tasks, such as a reading test, they select among the strategies they know. Strategies range from ineffective to effective:

(1) just read the questions, no need to read the passage;
(2) read the questions first, then skim the passage for a literal answer;
(3) select an answer that "looks good" to get through the dreaded reading test;
(4) "text matching" or matching words in the question stem or possible answers in a multiple choice question to words somewhere in the passage;
(5) choosing an incomplete answer by not considering all of the possible answers; and
(6) distinguishing between the "best" answer and a "less than best" answer.

EIa is concerned with the developmental nature of children's thinking and evaluating what they have been asked to read by their response to challenging questions. The EIa assumes part of the reason children score low is often a developmental combination of skimming or just reading the words of a passage, the question stem, and/or each foil, with limited thinking and evaluating (effort). This lack of effort is a symptom of low reading motivation as described in *Attribution Theory* (Bell-Gredler, 1986). Part of the task for educators is to motivate students and help them through this process of selecting and rejecting ineffective reading comprehension and answering strategies. Since children may use less efficient strategies based upon age and experience (Siegler, 2005), children with less literacy experience may utilize less efficient reading

comprehension answering strategies longer than their more literate counterparts. Teachers need to explicitly and consistently demonstrate "reading is thinking" and help children reject those less efficient strategies in favor of more efficient ones by providing students with more challenging and inferential activities. Beck, *et al*. (1997) found that inferential activities focusing on interpretation and intent became a powerful tool for changing students' approach to reading comprehension. When teachers and students are provided materials that require students to think about the author's intended message, beliefs, predictions and to evaluate how successful the author was at conveying his/her message, Beck found impressive changes in the classroom culture, with students more actively engaged in interactive discourse. The dynamic is self-reinforcing; as students and teachers engage in more thoughtful questioning (inference) they become more critical readers and thinkers.

EIa *with* Teachers

When reading comprehension errors occur, there may be no single cause for the error. Unlike the mathematical process, there is no one-to-one correspondence, no step-by-step process which leads the learner to the correct answer. The reading process closely resembles the thinking process in that there are multiple factors occurring simultaneously. When a student makes a mistake in comprehension it could be for a myriad of reasons, e.g. inadequacies in background knowledge, decoding ability, vocabulary, language structure, interest, motivation, attention to the task, etc. Therefore, as "outsiders" (outside the classroom) when a reading error is identified on reading assessments, it is not possible to absolutely determine the cause for the reading error "without asking the student" (Norris, et al, 2007), nor is it possible to tell the student how to go back through the learning sequence and fix the problem. To do this, we need a more informed, *internal view* of the learning environment *to compliment* the behaviors demonstrated by the student during diagnostic testing followed by remediation. With EIa reports and classroom teacher's knowledge of each student, teachers conference with students to determine what each student was thinking during the reading process. These conferences, called retrospective verbal protocols are "missed opportunities" (Gorin, 2007). With a combination of EIa reports and teacher conferencing using the EIa items as instructional aids, *teachers* determine each student's reading/thinking errors, and as a result, differentiate their current instruction.

Teachers need to explain the strategies and model how and when to use them. Following instruction, students would then need to practice and receive corrective feedback as to how successful they are in using the strategies and making inferences. But in order for teachers to determine if students are, in fact, using these strategies effectively, they need precise and timely information on what errors in thinking or errors in inference the student is making during the actual process of silent reading. Within a week of each EIa, teachers receive error reports and are told to keep the assessments and use them for instruction. Teachers treat these assessments as exercises and part of the delivered curriculum.

A Cognitive Model

In order to build a diagnostic assessment, a cognitive model is required. A cognitive model in educational measurement refers to a simplified description of human problem solving on standardized tasks at some convenient grain size or level of detail in order to facilitate explanation and prediction of students' performance, including their strengths and weaknesses. Assessments based on cognitive models of task performance should be developed so test items directly measure specific cognitive processes of increasing complexity in the understanding of a domain (Gierl, Leighton, & Hunka, 2007). EIa items were designed to identify the student errors and inefficient behaviors as defined in the following Cognitive Model.

EIa's Cognitive Model of Task Performance as Students Respond to Inferential Reading Comprehension Multiple Choice Test Items

Reading Comprehension is primarily measured by students responding to multiple choice test items. A reading comprehension test consists of one or more passages followed by mostly inferential, higher-order questions. The foils (wrong answers) are designed to distract the "weaker" readers. Sometimes there are only small differences between foils based upon nuances/inferences of one word/phrase, so students must read the question stem and each foil carefully and apply higher-order thinking in order to evaluate the differences between the "best" answer and a closely "related" foil. It is assumed that students come to the reading comprehension assessment task highly motivated and are able to carefully and thoughtfully read and understand the passages, questions and foils.

The **highest scoring** students will be motivated to accurately complete the reading task. They will expend time and effort by carefully reading each item stem and evaluating which of the presented foils is the "best". These students are also familiar with the testing format and the text of the passage and items are at their instructional or independent reading level. These students will also "look back" in the passage to re-read and clarify their perceptions and confirm their answers .

Slightly lower scoring students may be slightly less motivated and have obtained slightly lower reading levels. They will carefully read each item and consider the subtle nuances between the foils. However, these students will tend to develop more misconceptions than their high scoring peers and not be able to successfully discriminate between some of the foils that are closely related to the "best" answer. They will generally "look back" to re-read and clarify when they feel it is needed.

Still lower scoring students are less motivated, and will "look back" to clarify less often. These students may find reading (tests) less enjoyable and sometimes rush or not read the item stem as carefully as their higher scoring peers. They generally read each foil but may not carefully consider each foil. They tend to select the first foil that sounds good or select the more literal answer. Inference is more difficult and less utilized for these students than their higher scoring peers.

Lowest scoring students are the least motivated or they are reading passages/items that are approaching or at their frustration level. They may tend not to like reading and are prone to skimming and not "looking back" to re-read to clarify. Also, they may not read the item stem closely or may rely on their sometimes limited background knowledge to answer questions. They are most likely not to take the time to read all of the foils of an item and are unable or unwilling to evaluate the differences between foils. They look for literal answers. They feel selecting the right answer is a matter of luck, not effort. They may even decide not to complete all of the items of the Reading Comprehension test.

Reading Comprehension Errors Identified

EIa identifies three major types of silent reading comprehension errors and estimates the degree to which each student practices metacognitive strategies to clarify understanding of the text. Two of these errors were observed and reported by Pearson, *et al*., (1992): "answer grabbing" and "simple text matching." Psychometricians and item developers also utilize "text matching" when writing foils "distract" the poorer reader. Item developers also take a great deal of effort to carefully craft another type of distractor: what the authors call the "related" foil. Related foils are similar to the "most correct" answers but are not the "best" answers. Related foils may deviate from the "best" answer by subtle nuances or only a word or two and are designed to distract all test takers except the very highest scoring students. Students need to successfully discriminate between the inferential meaning of a related foil(s) and the "best" answer. The "related" foils are what some reading researchers (Drum, Calfee & Cook, 1980; Hill & Parry, 1988; Langer, 1987) have called "tricky" (Hiebert & Calfee, 1992). When examining the Federal and State mandated assessments in reading comprehension, many of the questions are "tricky".

In an effort to provide meaningful and timely information to classroom teachers, the three primary* types of errors identified by EIa are:

(1) **answer grabbing** — there is no support for this answer (foil) in the passage,
(2) **text matching**—the text was mentioned someplace in the passage but is clearly not the correct answer, and
(3) **related** foils—close to the correct answer but not the "best" answer.

The types of errors are identified and summed to help produce a quantitative error profile of each student. The teachers are asked to conference with students to confirm the identified types of misconceptions and inefficient behaviors. The qualitative confirmation of each student's pattern of weaknesses by teachers also allows parents and principals to understand and address the specific academic needs of students (as required by NCLB, Part A, Subpart 1, Sec 2221[b]3[C][xii], 2001).

* Other types of coded errors are: opposite (selecting a foil that is the opposite of what was asked); Looks like (selecting an answer that has many of the same letters as the answer); Pronoun referent (incorrect pronoun referent error) (also see Attachment A).

Each EIa foil is coded to identify a particular type of error. If the student selects a correct answer (coded C), then no error is recorded. EIa is designed to identify and report the errors each student is making and recommend the metacognitive strategies and behaviors that will help the reader overcome the errors. For example, in Exhibit 1, question 65, the student was told *where* to look back to clarify an answer (paragraph 4). Each foil is designed to elicit a particular type of common error that teachers often observe in their students. Errors are coded on the "Teacher Copy" of each EIa reading assessment.

Of the students who selected foil A (dentist), they probably re-called that *John Rock* was a dentist (in a different paragraph) but did not "look back" to paragraph 4 to clarify their answer. Foils A, B and D are coded "T" for "text matching errors". The "T" coded foils are summed across all reading items and passages. This sum becomes a measure of the number of errors caused by NOT "looking back" to clarify. If the student selected foil E, then s/he selected a foil that was not supported by the text. This student was "answer grabbing" (coded X). These answer grabbing foils are recorded and summed for a measure of answer grabbing errors. Students who skim the passage, or don't read the passage and just answer the questions based upon their background knowledge, tend to select these "answer grabbing" foils more than their better reading peers.

Exhibit 1 (Teacher Copy)

JOHN ROCK

Paragraph 4	At that time many of the laws were not fair to black people. John did not like these unfair laws. So he moved to Boston, Massachusetts to study law. John studied hard in law school, too. Later, he became one of the first black lawyers in Massachusetts.

Item 65 Why did the author write **paragraph 4**?

Error Codes	Foils	Comment
T	To show how he became a dentist.	(he was a "dentist" in paragraph 3)
T	To show John was a teacher.	(he was a "teacher" in paragraph 2)
C	To show John was a hard worker who wanted to help black people.	(inference)
T	To show John was the first black lawyer to be recognized by the Supreme Court.	(this was described in paragraph 5)
X	To show John moved to the South during the Civil War.	(he did not move to the South)

Exhibit 2 displays paragraph 4 from another passage titled *Climbing High*, with coded questions #65 through #68. Seven of the foils are coded as "related" (R). Related foils are good answers, just not the "best" answer. Students who select related foil(s) have generally demonstrated that they have read the passage with meaning, and are thinking about their answers. A student who answers "B" for #65 and #66 and "A" for #67 and #68 would receive a "low" scale score on traditional testing, but that student does have a good understanding of the passage and could discuss/debate concepts in the passage. However, in order to improve their *measured* reading achievement, this student will need to read with finer levels of discrimination and more carefully weigh all of the possible answers before responding to each question.

Exhibit 2 (Teacher Copy)

Story Introduction:

This is a story about Annie Peck; she was one of the first women to climb mountains.

CLIMBING HIGH

4 After several attempts, she became only the third woman ever to scale the Matterhorn Mountain in the Swiss Alps, and the first to make the climb in trousers rather than a cumbersome skirt. She continued climbing peaks throughout the world, achieving her greatest triumph with her record-breaking ascent of the north peak of Peru's 22,205 foot Mount Huascaran in 1908 at the age of fifty-eight! At the time, the exact height of Huascaran was not known and she believed she had climbed the highest summit in North or South America. In addition to the difficulty of the climb she had to cope with the ridicule and teasing of many male mountaineers, lack of money and equipment which, since it was designed for men, did often not fit her very well. For her ascent of Huascaran, she designed her own mountain shoes and had them made to her order. A few years later, Peru named part of the mountain after her.

65 Based upon the facts in paragraph 4, you can tell Annie enjoyed ...

X	A. being a good friend.	*(friends were not mentioned or inferred)*
R	B. doing crazy things.	*(some may believe this)*
C	C. mountain climbing.	*(implied, best answer)*
R	D. designing clothes and shoes.	*(she did do this)*
T	E. being teased.	*(simple text matching)*

66 From the statements in the passage, Annie would probably believe ...

T	A. all women should climb mountains.	*(not all & climb mountains = text matching)*
R	B. mountain climbing equipment is not designed for women.	*(until Annie designed the equipment)*
R	C. people should not tease others.	*(she was teased)*
C	D. women can do anything that men can do.	*(implied, best answer)*
T	E. women should visit New Hampshire.	*(simple text matching)*

67 Which statement **BEST** describes what the passage is **MOSTLY** about?

R	A. Annie was a woman who took many risks.	*(she did take risks)*
C	B. Annie was a person who wasn't afraid to follow her dreams.	*(implied, best answer)*
R	C. Women can be mountain climbers.	*(she did)*
T	D. Most mountains can be climbed.	*(text matching)*
T	E. Most people never climb mountains.	*(text matching)*

68 Annie went to college, became a mountain climber, designed her own mountain shoes and was always teased by boys. What does this tell you about Annie?

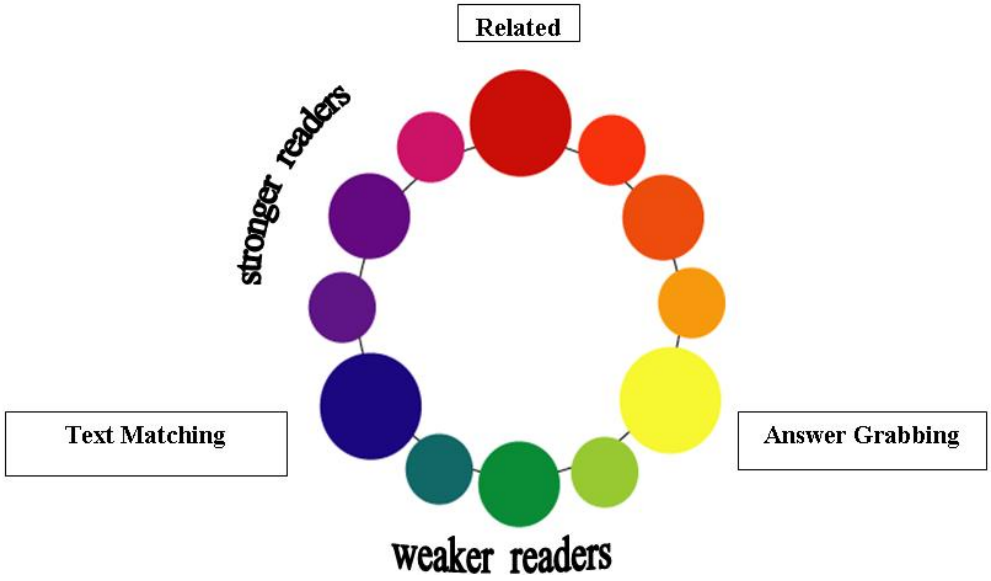
R	A. She didn't like boys.	*(never stated or implied)*
T	B. Annie learned Latin.	*(text matching)*
T	C. Annie could run fast too.	*(text matching, paragraph 1)*
T	D. Annie was a woman mountain climber	*(text matching)*
C	E. Besides being brave, Annie was a thinker too.	*(implied, best answer)*

EIa Error Reports in Reading

The EIa reading comprehension assessments tally and report each student's reading errors: answer grabbing, text matching and related. Since students use multiple strategies at the same time (Siegler, 2005; Pearson, Roehler, Dole & Duffy, 1992), the error profile of each student usually results in a mix of errors. The frequency of each type of error hints at a prognosis of what the teacher should do next. Exhibit 3, the color wheel, is a professional development conceptual aid, used to help communicate reading errors to teachers.

Exhibit 3

Identification of Reading Comprehension Errors



Preliminary research by the lead author seems to confirm the developmental progression that Siegler and Pearson, *et al*. have repeatedly reported. Poorer readers will select more answer grabbing foils (yellow). As students become better readers the frequency of answer grabbing will decrease and weak text matching (blue) errors will increase. Developmentally, yellow answer grabbing readers, will pass through green to blue. As students gain more experience and receive meaningful feedback, they become better readers; text matching errors will decrease and related errors will increase, passing through "purple" to red (see also EIa Reading Comprehension Error Reference Table, attachment A). The intention is not to color-code children, but rather to explain the complexity and developmental nature of EIa reading errors to better inform instruction.

There are three different types of EIa error reports in Reading:

1. Summary Error Report by student for each classroom (see Exhibit 4)
2. Detail Error Report: Foil selection by item for each classroom (see attachment B)
3. Individual Error Report for each student across all items to identify error patterns (see attachment C)

EIa reports are not based upon statistical models but on what is meaningful to classroom teachers with their current training, curricula and instructional program. Exhibit 4 is a sample EIa report of a grade 4 teacher's class. Based upon the student responses to 20 reading comprehension questions across two passages, each student's errors are summed and reported, usually within a week of taking the EIa. The variable "tot1" and "tot2" are the total correct (out of 10) for each of the two passages. An estimated performance level is provided using the terms with which teachers are familiar: Below Basic, Basic, Proficient, Goal and Advanced. The next variable (d_prof) provides an estimate of each student's "distance from proficiency." For teachers, this is a hopeful variable, indicating how many more points each student needs to be "proficient." The remaining variables (after correct) are the summed error categories as well as the number of questions each student "skipped" - a test taking avoidance behavior and motivation indicator.

Exhibit 4

Current Grade 4

Sample Summary Error Identification Report to a Teacher

tea906: story1: ChefSeattle_gr4 story2: ClimbingHigh_gr4 Report1: Error Summary

	tot1	tot2	rd_level	d_Prof	Correct	Related	Text Match	AnsGrab	Skip
JAMIE	10	10	Advanced	10	20	0	0	0	0
ZACHARY	8	10	Advanced	8	18	1	1	0	0
ANNA	6	8	Goal_Level	4	14	2	3	1	0
LUKE	4	9	Goal_Level	3	13	6	1	0	0
MILLIE	3	9	Proficient	2	12	3	3	2	0
YADI	3	7	Proficient	0	10	3	3	3	1
ALLEN	2	8	Proficient	0	10	1	8	1	0
DESHAUN	4	5	z_BasicLevel	-1	9	3	6	2	0
GUAD	3	5	z_BasicLevel	-2	8	5	5	0	0
SHELLY	2	5	zzBelowBasic	-3	7	4	4	5	0
KARL	6	0	zzBelowBasic	-4	6	1	2	1	10
AMAN	4	2	zzBelowBasic	-4	6	3	4	0	7
JOSE	2	3	zzBelowBasic	-5	5	7	7	1	0
ELLIE	2	3	zzBelowBasic	-5	5	3	3	9	0

Error Identification

Number Correct		Estimated Performance Level	Dist from Prof x2	Total correct out of 20	Types of Errors			
1st pass	2nd pass				related error	text match	answer grabbing	number of skipped items

In Exhibit 4, the lowest scoring student, Ellie selected nine (9) answer grabbing foils. Ellie should be asked why she selected each of these no-support answers. On the color wheel, she is more yellow than any other color, but she did also select three related and three text matching errors. The teacher has data to help guide student-teacher discussion, what some researchers call "penetration", (Gorin, 2007) and this error profile will probably be repeated in the future unless her behavior is modified with focused instruction.

Allen made primarily text matching errors. Although a typical error which seems easy to correct, it is a challenging task to get students to change their reading behavior without error identification and timely feedback. However, just knowing that his teacher is able to "see" his metacognitive errors, the student is likely to start practicing these previously taught fix-up strategies (Duel 1958, Hammond 1971; Meece, Blumenfeld, & Hoyle; 1988; Pressley, Woloshyn, *et al.* 1995).

Karl skipped 10 items (the second passage or half of the test). When asked by a teacher, Karl was surprised that his teacher knew he skipped any items because Karl *was never held accountable* for his reading behavior. He said: "I didn't feel like reading anymore". With EIa, students are now held accountable for their reading, rather than receiving no feedback from skipping items or selecting the first foil that looks good. Students who skip items on standardized tests will receive an underestimated scale score, but skipped EIa items are seen as exhibiting a symptom of low reading motivation or lack of "effort" according to Weiner's Attribution Theory (Bell-Gredler, 1986). This accountability and motivation issue needs to be immediately addressed with Karl and his unmotivated peers.

Jose selected seven related and seven text-matching errors. Although Jose only got 5 correct (out of 20), his selection of seven related foils demonstrates that he is reading and thinking about his reading, but lacks the experience in discriminating between a good answer and the "best" answer. Jose and many of his peers need practice in metacognitive fix-up strategies, as well as the dialogue that accompanies frequent rehearsal exercises in reading comprehension.

The Related Foil: Exercises in Critical Reading

The inferential and evaluative questions in Exhibit 1 and Exhibit 2 with related foils are difficult questions for our students who are usually exposed to much easier tasks. An unpublished study by the lead author found the assessments included in the district's purchased textbook/anthologies in grades 3 to 5 were comprised of only 10% inferential questions and 67% literal or factual questions, where the state and federal assessments are comprised of almost all inferential questions. In addition, the p-values of state and federal reading tests, according to other district researchers across the United States, are much lower than those in the purchased curriculum packages (.70 to .90), so students and teachers are generally exposed to relatively easy and/or literal items during the school year but are tested by the state with more difficult inferential items, often with closely related foils. A wide variety of researchers (Allington, 2001; Brown, 1991; Elmore, Peterson & McCarthey, 1996; Knapp, 1995; Tharp & Gallimore, 1989; Black & Wiliam, 1998) were reported by Applegate (2009) to have observed classrooms that do not engage readers in higher-order thinking, but emphasize literal recall. Both the purchased curriculum-based packages and teachers promote primarily open-ended questions. These open-ended questions, although an important pedagogy, are not anchored to a rigorous rubric; so many lower-level responses are acceptable to both teacher and student. As a result, both teachers and students do not have adequate opportunities to rehearse with more demanding multiple-choice questions which utilize difficult related foils and higher-order inference. Students often have difficulty with related foils.

One grade 4 teacher summarizes the difficulty of teaching students to evaluate the sometimes subtle differences between *related* foils and the "best" answer:

- As we are working on it [post-conference review if EIa items], often students will "get" it when it is a clear error. The related [versus the] correct answer does not come easily during the explanation.

The process of repeated teacher-student conferencing about the differences between EIa "related" versus "best" answers, with appropriate scaffolding, becomes a critical reading exercise at the same level of rigor as the state and federal high-stakes tests. Some students will learn and achieve, other students will not yet have the "ability" to comprehend the related/best nuances.

For many students, sometimes it seems there are two (or three) “BEST” answers, when, according to the item writer, there is one BEST answer with one or more related foils. These students do not, or cannot, comprehend the subtle nuances between the possible answers. The lack of “ability” to evaluate the differences between related and BEST foils may be the result of one or more of the following:

1. Reading motivation or lack of effort: it is much easier to select the first good answer,
2. Literacy or vocabulary development,
3. Previous learning experiences of primarily literal or easy answers to questions, and
4. Lack of assessment literacy or understanding what the assessment task requires, some populations may require “a more comprehensive orientation to the testing process” (AERA, 1999; p 61).

As difficult as these related foils are to evaluate (and teach), the successful evaluator of these foils, demonstrates the reader’s ability to read the vocabulary of the passage, item stems and foils with a high degree of inferential understanding. Many students will not demonstrate this pattern of success; others may, with instruction guided by diagnostic information. If the purpose of a test is to determine “ability estimates,” then the related foil is an important distractor. However, it is a fair and recommended ethical practice (AERA, 1999; p 61), that test takers are aware of this esoteric testing format.

Reading Comprehension Support Exercises

Because of the weaker curriculum assessments that teachers and students are repeatedly exposed, teachers need additional instructional materials, at the same challenging levels of difficulty and discrimination as the state assessments to support the delivered reading curriculum. The RAND review concludes that teachers who give students challenging tasks, and opportunities for collaborative learning increase their motivation to comprehend text (National Research Council, 2004). As a result of the relatively weak base-curriculum packaged assessments and challenging state assessments, this district has begun to develop its own additional instructional materials, at this same level of rigor as the state assessments to support the delivered reading curriculum.

As a necessary complement to small and whole group discussion/debate using EIa assessments and reports, a series of challenging, weekly Reading Comprehension Support Exercises (RCSE) are being developed to continue the dialogue. Many of the Reading Comprehension Support Exercises are based upon the district's core anthology. Other exercises are unconnected to the anthology, resembling the "cold" reading selections similar to state and federal tests. The exercises consist of a very short passage with one inferential question and five multiple-choice foils consisting of one "best" answer and at least one or more related foils. Reviewing student responses create a positive dialogue where new understanding between the student and a more knowledgeable adult are co-constructed (Wells, 2000). Over time, the Reading Comprehension Support Exercises Item Bank, consisting of hundreds of these rehearsal exercises across grades 2 to 9, will be available for each teacher's use during small and whole group instruction. RCSEs are expected to be teacher selected as needed, and perhaps used as the "reading comprehension question of the day". An example of an RCSE is provided in Exhibit 5 which was developed from a revised newspaper article.

Teachers are asked to pass out the student copy of the Reading Comprehension Support Exercise during small group work. The small group discussions would follow this suggested sequence:

(1) Discuss **"***What is this question really asking me?"*
(2) Have students discuss their answers and locate the evidence (explicit or implied) within the brief passage that supports their "best" answer.
(3) Guide the discussion ending with what the item writer believes is the "best" answer and *why* the related foils are not considered the "best" answer.

During this sequence, students gain experience rehearsing reading comprehension of the item stem, and the nuances of each item distractor/foil. Both teachers and students are learning what is expected of them on standardized state and federal reading tests. These EIa and RCSE instructional activities attempt to provide what Popham (2008) calls essentials of assessment *for* learning.

Exhibit 5 (Teacher Copy)

Reading Comprehension Support Exercise

Salute to the Irish Brigade*

Last week, despite the inclement weather, patriots turned out in numbers recently for the wreath-laying ceremonies at the Ninth Regiment Civil War monument in the Hill section of the city. They paid tribute to the brave soldiers who lost their lives fighting for the "Ninth," made up mainly of Connecticut Irishmen, during the Civil War.

The Second Company Governor's Foot Guard Fife and Drum Corps provided music. A piper from the Gaelic Highland Pipe Band played the bagpipes while a New Haven police officer and bugler, stood ready.

A reception was held following the ceremony at the New Haven Gaelic Club in East Haven.

In the author's "Salute to the Irish Brigade," who were the "patriots" who went outside in bad weather for the wreath-laying ceremonies?

Text Matching	*A.*	The Ninth Regiment of the Civil War
Text Matching	*B.*	Connecticut Irishmen during the Civil War
Related	C.	The Second Company Governor's Foot Guard
Text Matching	*D.*	The New Haven Gaelic Club in East Haven
Correct	*E.*	The people watching and participating in the activities at the Hill section of the city last week.

* Revised with permission from: Salute to the Irish Brigade. (2006, November 26). *New Haven Register*. F2.

Scaffold Instruction: What To Do Next

The primary difference between traditional state and national achievement tests and EIa is the former are *designed* to spread test takers apart with scale scores, and the latter is *designed* to help identify patterns of student errors in thinking/strategy use. Another major difference between EIa and summative state tests is that the actual EIa items are discussed and debated with students the next day or two after EIa administration. Teachers are encouraged to "use the EIa for instruction" while students can still remember their thoughts and responses. This is where teachers may also use "think alouds" (Davey, 1983) and "retroactive verbal reports" (Leighton & Gierl, 2007). Just like RCSE, students are questioned about *why* they answered as they did: Where is the support for your answer? What evidence did you use from the text? Does it make sense? Teachers and students also debrief challenging multiple-choice items and focus on the small nuances between related and "best" answers, as well as emphasizing the fix-up reading strategies. The dialogue generated from the EIa questions and answers becomes another opportunity for progressive discourse concerning the teaching of critical thinking and understanding of ideas in the text (Wells, 2000; Block, Grambrell & Pressley, 2002).

A scaffold in the context of teaching is the support teachers provide students to help them move to a higher level of learning or cognitive understanding. Hogan and Pressley (1997) suggest that active diagnosis of student learning is a defining element of scaffolded instruction. They also maintain that a commitment to letting students do the thinking or at least, share in the thinking, and the teacher's gradual removal of support are essential elements of scaffolding. Specially crafted error identification tasks assist teachers in scaffolding by providing teachers with the incremental information they need to make decisions as to where the students are in their developmental levels of reasoning and learning.

Aside from whole group and small group discussion and debate, students need continued opportunities for further practice in strategy use, under the guidance of the teacher. Guided reading groups provide time for teachers to review and positively reinforce reading strategies and metacognitive "fix-up" strategies such as: looking-back, re-reading, chunking, reading ahead, adjusting rate of reading and comprehension monitoring. A gradual release of responsibility continues during student independent reading, as students practice the strategies on their own. For the teacher, this is the perfect time to conference with students about independent strategy use. In this way, the teacher is able to scaffold instruction at every point in the learning sequence:

The Gradual Release of Responsibility

To (1) instruction *to* the students by the teacher,
With (2) instruction and practice *with* students and teachers together, and
By (3) students practice *by* themselves, with the teacher observing and noting where continued instruction needs to take place.

EIa (and RCSE) provides a missing element in the gradual release of responsibility because it gives both teachers and students the precise feedback needed to determine the degree of strategy acquisition at every step of learning. This makes it is a dynamic assessment providing critical information which reflects the changes in each student's cognitive and affective development as student learning evolves.

The flexible grouping of students for instruction is based upon the metacognitive errors each student demonstrates, rather than a normative scale score. Guided by the EIa reports, Exhibit 6, compiled by the second author, provides recommended activities and strategies to address student metacognitive errors.

Exhibit 6	
Student Errors	**Suggested Strategies and Activities**
Answer Grabbing Errors	Possible random response/guessing answer grabbing/no support in passage. Student selection of foils without text support, indicates students may not have read the passage or passage is above reading level. Student needs to be coached in "fix-up" strategies, encouraged to re-read the passage and apply reading strategies indicated below.
Text Matching Errors	Have student practice locating facts in small groups and rehearse fix-up strategies such as "look back" (Garner, 1987) to find support in the passage, Directed Reading-Thinking Activity (Stauffer, 1969), Questions Answer Relationship (Raphael, 1986), and K-W-L (Ogle, *et al*., 1986). This means demonstrating for the whole group and the forming of small groups where students read passages, answer questions and find evidence in the passage to support their answers (re-reading activities).
Related Errors	There is more than one good answer but the question asks for the "best" correct answer. This student did not select the "best" inference. Demonstrate "Think-Alouds" (Davey, 1983) to whole group and allow student to rehearse in small groups. Also, use "Questioning the Author" (Beck, *et al*., 1997), think-along (Duffy, *et al*., 1988) and "Reciprocal Teaching" or "I Wonder Why" (Palicsar, *et al*., 1984), QAR, DRTA or higher-order questioning. Also select from the Reading Comprehension Support Exercises Item Bank for additional inferential materials to use during small group and/or one-to-one conferencing with the student.

Partial Qualitative Results

As a result of NCLB, this district's state department of education has required an external evaluation group to conduct critical in-depth evaluations of districts "not making AYP" and to produce a formal evaluation report. One section of this report on "Assessments Aligned with Curriculum and Instruction" states that the district's Error Identification assessments

> "… is a useful data tool developed in the district to provide interim assessments in reading and mathematics three times a year in grades 2 though 9. It is aligned with the [state testing] and provides a means of error analysis which teachers and administrators find helpful. One of its key strengths is that it provides rapid feedback on test scores which teachers can use readily to gauge what has been learned and what needs further work …Teachers like using the [EIa] because it provides timely and detailed information regarding weaknesses in learning, providing an 'error analysis' of test strands. The data teams in some of the schools are using the error analysis well with scrutiny of students' work to help target their instruction to address weak areas. However, there is more to do to ensure the teachers are using the data routinely to differentiate the instruction and match it to levels of ability in the class" (Wheatley, 2008).

Anecdotally, the EIa reports have also been utilized in PPT's and parent conferences to help explain student needs: cognitive processes, misconceptions and error patterns. EIa reading data has been utilized to help assess if students are, or are not, eligible for special education services. These assessments "are valued by teachers and parents" (Wheatley, 2008).

How Teachers Use EIa and Reports

In order to help determine how teachers and students actually use these diagnostic assessments and reports, (with very little staff development) a voluntary survey was administered to all teachers of language arts teachers in grades 2 to 9. About 15 % of the reading, language arts, English, Special Education and ELL teachers returned this voluntary utilization survey. The teacher comments are grouped by how EIa clarifies student thinking/behaviors, helps teachers plan activities, and how EIa helps students become involved in their learning. One teacher included many of these attributes in his/her summary of the diagnostic benefits of EIa and EIa reports:

> This gives me an in depth look in an easy to read and quick document. I've tried very hard to confer with students individually to discuss their errors and set goals for improvement. I think they benefit from knowing how to improve with specific goals. These concepts help me identify problems/concerns early on in the year. This gives me more time to help students improve and become better readers. (classroom teacher, grade 4).

Many of the teacher responses commented on student thinking:

- I use them to see what they are thinking … which errors seem to be used consistently-what they are "tricked on". I turn it into a game-type activity-student vs. teacher. (reading teacher, grade 2-5).
- If students are aware of the errors they make, determined students who want to do well, will change their behavior both in reading and testing (classroom teacher grade 5).
- Looking for patterns and trends certainly helps teachers to understand students' metacognitive errors (reading teacher, grades 3-4).
- The error identification gives insight on their thought processes (classroom teacher, unknown grade).
- The foils help me to identify how students are thinking through the question (classroom teacher, grade 2).
- They now don't grab the first answer that they connect to. They take more time to evaluate and critique each choice (reading teacher, unknown grade).

Other teacher comments emphasized the diagnostic aspects of EIa which help teachers plan activities.

- Students are alerted to when they need to go back and reread to find the answer. We can tell which students need help going back to reread, which students are close, which students are just picking answers (classroom teacher, grade 3).

- I meet with students in small groups or individually to discuss the test. We talk about how and why they selected their answers. I use the spreadsheets to pin-point which skills need more review and reinforcement. This helps drive my instruction (classroom teacher, grade 4).

- Related errors help me determine what students are reading well but not carefully enough. No-support errors identify students who clearly can't read on grade level and/or aren't taking it seriously (classroom teacher, grade 4).

- Keeping the test booklets and handing them back to students is helpful in going over errors they might have made. I especially find it useful in discussing why they chose their answers and it makes them accountable for their choices (classroom teacher, grade 4).

- The results from these tests allow teachers to identify their students' needs and plan instruction to meet those needs" (special education teacher, grades 3-4).

- We revisit the text and go over all answers orally. Students will share their thinking strategies, find where in the text they found the answer and explain why an answer was incorrect. I try to do this within one week after the testing (classroom teacher grade 6).

- Conference with them to address these mistakes and to review with parents (classroom teacher, grade 4).

- We help them understand their identified errors and model how to change their reading behaviors by modeling in small groups or on the overhead (reading teacher, unknown grade).

- During these exercises it is easy to see which students get it and which don't. Plus when students are sitting there in small groups we can ask "what were you thinking?" (classroom teacher, grade 5).

- Looking for patterns in foil choices (individual or whole class) helps me to target specific question styles and strategies (reading teacher, grade 3- 4).

Some teacher comments explained how students benefit from EIa and the difficult task of teaching reading:

- Now I feel confident explaining the types of answers to students in terms they can understand (classroom teacher, grade 6).
- Students understand that they need to re-read and find supporting text evidence to get the best responses (reading teacher, unknown grade).
- When students know the exact reasons they are making errors, they are able to identify and correct them more easily (classroom teacher, grade 9).
- We call them tricky questions for M.C.-students are excited to "not get tricked" (reading teacher, grade 4).
- Practice over time helps students feel more confident in their ability to understand what is being asked and to become more proficient (reading teacher, unknown grade).
- Some students understand but the students who have the most difficulty just don't get it or can't remember to change their behaviors (reading teacher, grade 7).
- Understanding their misconceptions is one thing - getting them to change the misconception is the difficult part (classroom teacher, grade 7).

How Other LEA's Use EIa

Two other LEA's have been using these same error identification assessments, software and reports, one small district in the same state and one large district from a different state. The following summary statements describe how the smaller LEA use EIa:

> As a coach, these tests provided useful data for the teachers. I was able to guide them in forming differentiated groups to use during their readers and writers workshop. The data also guided us on how to plan for intervention instruction. We were able to diagnose specific weaknesses and find materials for differentiated instruction.
>
> These diagnostics were vital to our intervention model and helped to support the teachers, tutors and students. Overall by using the tests for instruction and sharing results with the students they took ownership of their scores and realized that a test can be useful and have meaning for their learning.

Partial Quantitative Results

Each year, state mandated tests are administered in March in grades 3 to 8. The state's Fourth Generation tests were first administered in March of 2006. The second administration across grades 3 to 8 were administered in March 2007. Using the NCLB, AYP evaluation model the percent of students above "proficiency" in mathematics "increased" (6% to 12%) across all grades. State test results using the NCLB, AYP evaluation model in reading showed mixed results, but the development of the more complex EIa reading assessments are at least one year behind the math assessments and reports. The unmatched grade 3 cohort of 2006 to grade 5 of 2008, reported "increases" of about 5% per year (45% proficient in 2006, 50% proficient in 2007 and 54% proficient in 2008 grade 5). No change reported for the grade 6 to 8 cohort.

The state has developed Vertical Scale Scores in reading. The matched cohort from grade 3 in 2006 to grade 5 in 2008 recorded greater than average growth (57 Vertical Scale Score point gain versus the state average of 53). The matched cohort from grade 6 in 2006 to grade 8 in 2008 recorded only one (1) Vertical Score point over the state average growth (27 versus 26).

Conclusion

For the past 100 years, assessments have been designed by professional test developers for decision makers who require summative evaluations using ability estimates across tens of thousands of students at each grade level. In an effort to be instructionally informative *for the individual teacher*, a new generation of assessments is being developed. The diagnostics of these new assessments are based upon the identification of each student's misconception, inefficient strategies and/or pattern of errors. EIa is a multi-year attempt to operationalize an error identification system that goes beyond the laboratory and selected classes to a district-wide initiative. The primary purpose of EIa is to identify and promptly report each student's errors in thinking and metacognitive strategy use so teachers can confirm and guide remediation. The errors identified by the assessments are the same errors that students make during regular school work, as well as high-stakes testing. The assessments are dynamic in nature because they offer insights into student thinking as students engage in the act of reading or computing mathematics in grades 2 to 9. In this district, EIa represents a system that integrates both instruction and diagnostic assessments both teachers and parents value.

Public school districts are large, complex organizations with multiple initiatives and interactions between initiatives. Implementation of any one initiative is often difficult to accurately access. However, preliminary results indicate teachers value the EIa and subsequent error reports may have helped raise test scores across the district in mathematics. Error identification in reading is more complex, because there are multiple causes for poor reading comprehension. For this reason, direct instruction in metacognitive strategies may not immediately impact measured reading comprehension. For both mathematics and reading we are awaiting the 2009 state mandated test results and future evaluation results. However, by informing instruction for teachers, and providing insights for students into their own learning, Error Identification assessments begin a new district level genre of instructionally relevant assessments serving both teachers and students.

Attachment A

Districtwide Assessments: Reading Comprehension Error Reference Table

			← Weaker →						
							← Stronger →		
Item #	Question Type	Strand #	No Support	Looks like "*letter matching*" *(vocabulary)*	Opposite	Anaphoric Pronoun Referent Error	Text Matching or Weak Support	Related Support	Correct or Best Answer
1	Factual	(1) Initial Understanding	**X**		**O**		**T**		C
2	Factual	(1) Initial Understanding	**X**		**O**		**T**		C
3	Vocabulary	(1) Initial Understanding	**X**	**L**	**O**				C
4	Vocabulary	(1) Initial Understanding	**X**	**L**	**O**				C
5	Inference	(2) Develop an Interpretation	**X**		**O**	**A**	**T**	**R**	**C**
6	Inference	(2) Develop an Interpretation	**X**		**O**	**A**	**T**	**R**	**C**
7	Evaluative	(1) Initial Understanding	**X**				**T**	**R**	**C**
8	Inference/ evaluative	(2) Develop an Interpretation	**X**		**O**		**T**		**C**
9	Evaluative/ Prediction/ Simile/ Metaphor	(4) Exam Content and Structure	**X**			**A**	**T**	**R**	**C**
10	Evaluative/ Prediction/ Simile/ Metaphor	(4) Exam Content and Structure	**X**			**A**	**T**	**R**	**C**
			← **metacognitive errors** →						

sch: WMS tea_read: M Jane Detail_Report1: for GivingThanks_g7

id	lname	fname	rdcomp_level	totr awg7	tot1	Q51F z	Q52F z	Q53V z	Q54V z	Q55I z	Q56I z	Q57E z	Q58M z	Q59E z	Q60E
5010	I	Jame	5Advanced	18	8				b T						a T
10333	F	Roby	5Advanced	18	10										
2526	I	Eric	5Advanced	17	9						a T				
3004	B	Zach	5Advanced	17	9						e T				
2433	B	Jama	4Goal	16	9						e T				
4184	J	Garr	4Goal	16	8				d X		a T				
4803	H	Kath	4Goal	14	8						a T		c R		
1459	S	Ashl	3Proficient	13	5		b T	a T	a L				c R		d T
4132	G	Brit	3Proficient	13	5	c R		d L	e O	d T			c R		
4595	C	Bian	3Proficient	13	8						a T		c R		
4688	R	Deni	3Proficient	12	6				b T		a T		c R		d T
7949	D	Cole	3Proficient	12	5			a T	b T		a T		c R	d R	
4626	D	Chai	2_Basic	11	6				e O		a T		c R		a T
13523	P	Mich	2_Basic	11	5					d T	e T		c R	c T	d T
13644	V	Jose	2_Basic	11	6				e O		a T		c R		b R
22054	D	Ana	2_Basic	11	6						a T		c R	d R	a T
6049	J	Kaly	2_Basic	10	4			a T	b T		e T		c R	d R	a T
12098	R	Norm	1b_BBasic	9	5				d X		e T		c R	d R	a T
22050	A	Heri	1b_BBasic	8	4			c X	e O		e T		a T	c T	b R
8581	P	Shei	1a_BBasic	5	3			c X	a L	c T	e T	e R	b T	d R	
7761	A	Rich	1a_BBasic	3	2	a T	b T		e O	b X	e T	e R	b T	d R	

Number of cases read: 21 Number of cases listed: 21

Attachment B

Meriden's Cognitive Error Description in Reading: Individual Report (gr 7)
January 2008 Grade 7

School: L Report for: 134 Michael teacher: R , Myr

Passage 1: GivingThanks_g7	Passage 2: TheBullfight_g7	Passage 3: Editing/Revising
#51:	#61:	#71:
#52:	#62:	#72: a tense
#53:	#63: a R	#73:
#54: b T	#64:	#74: e colon
#55: d T	#65: c R	#75:
#56: e T	#66:	#76:
#57:	#67:	#77:
#58: c R	#68:	#78:
#59:	#69:	#79:
#60: a T	#70: e R	#80:

Total Reading Comp of 20: 12 MDA Reading Comprehension Level: 3Proficient

Total Edit/Revising of 10: 8

Total Overall Reading of 30: 20 MDA Overall Reading Level: 2_Basic

Fall DRP Unit Score: 57 Fall DRP Performance Level: 4_Goal

Type of Reading Error Summary:

Number of items correct out of 20:	12
Number of Related foils selected:	4
Number of NOT Looking Back foils selected:	4
Number of Answer Grabbing foils selected:	0
Number items Skipped out of 20:	0

Description of Error Codes for Passage 1 and Passage 2

Error Code	Description
#	The item number.
a, b, c, d, e	The foil selected (a, b, c, d, or e)
R	Related foil, a good answer but not the BEST answer.
T	Text Matching, this foil was somewhwere in the passage, but it is not close to the correct answer.
O	Opposite from the correct answer.
X	These "Answer Grabbing" foils are not in the passage, or inferred, but they sound like a reasonable answer if you did not read the passage.

References

Allington, R.I. (2001). *What really matters for struggling readers: Designing research-based programs*. New York: Addison-Wesley Higher Education.

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

Applegate, A.J., Applegate, M.D., McGeeha, C. M., Pinto, C. M., & Kong, Ailing (2009). The Assessment of thoughtful literacy in NAEP: Why the states aren't measuring up. *The Reading Teacher* Vol. 65 (5), 372-381.

Baker, L., & Brown, A.L. (1984). Metacognitive skills and reading. In P.D. Pearson, (ED), *Handbook on reading research.* (pp. 353-394). New York; Longman.

Beck, K.L., McKeown, M.G., Hamilton, R.L., & Kucan, L. (1997). *Questioning the author: An approach for enhancing student engagement with text.* Newark, DE: International Reading Association.

Bell-Gredler, Margaret E., (1986). *Learning and instruction, theory into practice*. (pp. 274-313). New York; Macmillan Publishing Company.

Bennett, R.E., (1993). Toward Intelligent Assessment: An Integration of Constructed-response Testing, Artificial Intelligence, and Model-Based Measurement. In Frederiksen, N., Mislevy, R.J., & Bejar, I. (Eds). (1993). *Test theory for a new generation of tests*. (pp. 104-115). Hillsdale, N. J. LawrenceErlbaum Associates.

Black, P., & Wiliam, D. (1998). Assessment and classroom learning. *Assessment in Education.* 5(1), 7-74.

Bloom B.S. (1976). *Human characteristics and school learning.* New York*:* McGraw-Hill Book Co.

Bloom B.S. (1980). The new direction in educational research: Alterable variables. *Phi Delta Kappan*, February, 384.

Block, Gambrell, Pressley, (2002). *Improving comprehension instruction.* (pp. 34, 395). Newark, DE: International Reading Association.

Brown, R.G. (1991). *Schools of thought: How the politics of literacy shape thinking in the classroom.* San Francisco: Jossey-Bass.

Cronbach, L. J. (1977). *Educational psychology (3rd edition).* New York: Harcourt, Brace, Jovanovich, Inc.

Davey, B. (1983). Think-aloud-modeling the cognitive processes of reading comprehension. *Journal of Reading*, 27, 44-47.

Downing, S., & Haladyna, T. M. (Eds). (2006). Handbook of test development. Mahwah, N.J. Lawrence Erblaum Associates.

Drum, P.A., Calfee, R. C., & Cook, L. K. (1980). The effects of surface structure variables on performance in reading comprehension. Reading Research Quarterly, 16, 486-514.

Duel, H. J. (1958). Effect of periodic self-evaluation on student achievement. *Journal of Educational Psychology*. 49, 197-199.

Duffy, G.G.I.R. Roehler and B.A. Herrmann (April 1988). Modeling mental processes helps poor readers become strategic readers. *The Reading Teacher* 41, 762-767.

Elmore, R.F., Peterson, P.L., & McCarthey, S.J. (1996). *Restructuring in the classroom: Teaching, learning, and school organization*. San Francisco: Jossey-Bass.

Embretson, S. (1993). Psychometric models for learning and cognitive processes. In Frederiksen, N., Mislevy, R.J., & Bejar, I. (Eds). Test theory for a new generation of tests. (p. 125). Hillsdale, N.J. Lawrence Erlbaum Associates.

Feuer, M.J., Holland, P.W., Green, B.F., Bertenthal, M.W., & Hemphill, F.C. (Eds.) (1999). *Uncommon measures: Equivalence and linkage among educational tests.* Washington, DC: National Research Council.

Flavell, J. H,. (1978). Metacognitive development. In J. M. Scandura & C. J. Brainerd (Eds), *Structural process theories of complex human behavior* (pp. 213-245*).* Ayphen & Rijn, The Netherlands: Sijtoff & Noordhoff.

Garner, R., Macready, G.B., & Wagoner, S. (1984). Readers' acquisition of the components of the text-look back strategy. *Journal of Educational Psychology*, 76, 300-309.

Garner, R. (1987). *Metacognition and reading comprehension*. Norwood, NJ: Ablex.

Gierl, M. J. & Leighton, J. P. (2007). Directions for future research in cognitive diagnostic assessment. In J. P. Leighton & M.J. Gierl (Eds.), *Cognitive diagnostic assessment for education theory and applications*. New York, NY. Cambridge University Press.

Gierl, M. J., Leighton, J. P. & Hunka, S. M. (2007). Using the attribute hierarchy method to make diagnostic inferences about examinees' cognitive skills. In J. P. Leighton & M.J. Gierl (Eds.), *Cognitive diagnostic assessment for education theory and applications*. New York, NY. Cambridge University Press.

Gorin, J. S. (2007. Test construction and diagnostic testing. In J. P. Leighton & M.J. Gierl (Eds.), *Cognitive diagnostic assessment for education theory and applications*. New York, NY. Cambridge University Press.

Haladyna, T.M. (2004). *Developing and validating multiple-choice test items*. (p 265). Mahwah, N.J. Lawrence Erblaum Associates.

Haladyna, T.M. (1997). *Writing test items to evaluate higher order thinking*. Boston, MA. Allyn and Bacon.

Hammond, K. R. (1971). Computer graphics as an aid to learning. *Science*. 172, 903-908.

Hiebert and Calfee (1992). Assessing literacy. In Samuels, S.J. & Farstrup, A.E. (Eds), *What research has to say about reading instruction.* (p 73). Newark, Delaware: International Reading Association.

Hill, C., & Parry, K. (1988). *Reading assessment: Autonomous and pragmatic models of literacy* (Literacy Center Report 88-2). New York: Teachers College Press.

Hogan, K., & Pressley, M., (1997). *Scaffolding student learning.* (p. 190). Louiseville, Quebec: Brookline Books, Inc.

Huff, K & Goodman, D.P. (2007). The demand for cognitive diagnostic assessment. In J. P. Leighton & M.J. Gierl (Eds.), *Cognitive diagnostic assessment for education theory and applications*. New York, NY. Cambridge University Press.

Jacobs, J. & Paris, S. (1987). Children's metacognition about reading: issues in definition, measurement, and instruction. *Educational Psychologist*, 22, 255-278

Kelleher, M. E. (1984), *Higher-order learning through higher-order feedback: An orientation program for freshman at MIU*. Unpublished Doctoral Dissertation, University of Cincinnati, Cincinnati, OH

Knapp, M.S. (1995). *Teaching for meaning in high-poverty classrooms.* New York:Teachers College Press.

Kolen, M.J. & Brennan, R.L. (2004). *Test equating, scaling & linking, methods and practices (2nd edition).* New York: Springer.

Langer, J. A. (1987). The construction of meaning and the assessment of comprehension: An analysis of reader performance on standardized test items. In R.O. Freedle & R. P. Duran (Eds.), *Cognitive and linguistic analyses of test performance*. Norwood, NJ; Ablex.

Leighton, J. P. & Gierl, M. J. (2007). Verbal reports as data for cognitive diagnostic assessment. In J. P. Leighton & M.J. Gierl (Eds.), *Cognitive diagnostic assessment for education theory and applications*. New York, NY. Cambridge University Press.

Luecht, R. M. (2007). Using information from multiple-choice distractors to enhance cognitive-diagnostic score reporting. In J. P. Leighton & M.J. Gierl (Eds.), *Cognitive diagnostic assessment for education theory and applications*. New York, NY. Cambridge University Press.

Mestre, J.P., Gerace, W.J., Dufresne, R.J., & Leonard, W.J. (1997). Promoting active learning in large classes using a classroom communication system. *In the changing role of physics departments in modern universities: Proceedings of ICUPE;* E. F. Redish & J. S. Rigden (Eds). (p 1021). The American Institute of Physics.

Mayo, K. E. (1992). Learning strategy instruction: Exploring the potential of megacognition. *Educational Psychologist*, 92, 130-134.

Meece, J. L., Blumenfeld, P. C., & Hoyle, R. H. (1988). Student's goal orientations and cognitive engagement in classroom activities. *Journal of Educational Psychology*, 80, 514-523.

McKeachie, W. J. (1974). The decline and fall of the laws of learning. *Educational Research*, March, 7-11.

Mislevy, R.J. (1993). Foundations of a new test theory. In Frederiksen, N., Mislevy, R.J., & Bejar, I. (Eds). *Test theory for a new generation of tests*. (p. 32). Hillsdale, N. J. Lawrence Erlbaum Associates.

National Research Council (2004). *Learning and instruction.* A SERP Research Agenda. (pp.54-56). Washington, D.C.: The National Academies Press.

Norris, S. P., Macnab, J. S. & Phillips, L. M. (2007). In J. P. Leighton & M.J. Gierl (Eds.), *Cognitive diagnostic assessment for education theory and applications*. New York, NY. Cambridge University Press pp 82-83.

Ogle, D. (1986). K-W-L; A teaching model that develops active reading of expository text. *The Reading Teacher*, 39, 564-570.

Osterlind, S.J. (1998). *Constructing test items: multiple-choice, constructed response, performance, and other formats*. Boston. Kluwer Academic Publishers.

Palincsar, A.S., & Brown, A.L. (1984). Reciprocal teaching of comprehension-fostering and monitoring activities. *Cognition and Instruction,* 1, 117-175.

Pearson, P.D., Roehler, L.R., Dole, J.A. & Duffy, G.R., (1992). Developing expertise in reading comprehension. In Samuels, S.Jay & Fastrup, Alan E. *What research has to say about readinginstruction,* (second edition). (pp157-168) International Reading Association.

Popham, J. W., (2008). Classroom assessment *what teachers need to know*. Pearson Education, Inc. 272-273

Pressley, M., Goodchild, F., Fleet, J., Zajchowski, R. & Evans, E.D. (1989). The challenges of classroom strategy instruction. *Elementary School Journal,* 89, 301-342.

Pressley, M., Woloshyn, V. and Associates (1995). *Cognitive strategy instruction that really improves children's academic performance,* (second edition). (pp. 74-75.) Brookline Books, Cambridge, MA.

Raphael, T. E., & Pearson, P.D. (1985). Increasing students' awareness of sources of information for answering questions. *American Educational Research Journal,* 22, 217-236.

Raphael, T. E. (1986). Teaching children Question-Answer Relationships, revisited. *The Reading Teacher,* 39, 516-522.

Siegler, R. S., (2005). American Psychologist, Children's Learning. 60 (8), 769-778.

Snow, C. E., Chair, (2002). Reading for understanding, toward a R&D program in reading comprehension. RAND Reading Study Group. (pp. 53-55). Santa Monica: RAND.

Snow, R.E. & Lohman, D.F. (1993). Cognitive psychology, new test design, and new test theory: An introduction. In Frederiksen, N., Mislevy, R.J., & Bejar, I. (Eds). *Test theory for a new generation of tests*. (pp. 7-11). Hillsdale, N. J. Lawrence Erlbaum Associates.

Stauffer, R. (1969). *Directing the reading-thinking process*. New York: Harper & Row.

Tharp, R. G., & Gallimore, R. (1998). Rousing schools to life. *American Educator,* 13(2), 20-25, 46-52.

Thorndike, E. L. (1931). *Human Learning.* (p.9). New York, Century.

Wells, G. (2000). Dialogic Inquiring in Education. In Lee, C.D. and Smagorinsky, P. (Eds), *Vygotskian perspectives on literacy research.* (pp. 71-73*)*. New York: Cambridge University Press.

Wheatley, D., December (2008) Quality review report for meriden public schools. (p 4,10).Cambridge Education (LLC) for the *Connecticut State Department of Education*.

The co-authors are Educational Psychologists with backgrounds in learning theory.

Alvin F. Larson, Ph.D., Director Research and Evaluation,
Meriden Public Schools, 22 Liberty Street, Meriden, Connecticut, USA.
Dr. Larson's area of concentration is Evaluation and Measurement;
email: al.larson@meriden.k12.ct.us.

Maureen E. Kelleher, Ph.D., Director of Language Arts, PreK-5,
Meriden Public Schools, 22 Liberty Street, Meriden, Connecticut, USA.
Dr. Kelleher's concentration is Reading and Human Development;
email: maureen.kelleher@meriden.k12.ct.us