Exploring Dynamic Assessment as a Means of Identifying Children At Risk of Developing Comprehension Difficulties
Amy M. Elleman, Donald L. Compton, Douglas Fuchs, Lynn S. Fuchs and Bobette Bouton

J Learn Disabil 2011 44: 348 originally published online 17 June 2011
DOI: 10.1177/0022219411407865

The online version of this article can be found at:
http://ldx.sagepub.com/content/44/4/348

Published by:
Hammill Institute on Disabilities

and
SAGE

http://www.sagepublications.com

Additional services and information for Journal of Learning Disabilities can be found at:

Email Alerts: http://ldx.sagepub.com/cgi/alerts

Subscriptions: http://ldx.sagepub.com/subscriptions

Reprints: http://www.sagepub.com/journalsReprints.nav

Permissions: http://www.sagepub.com/journalsPermissions.nav

Citations: http://ldx.sagepub.com/content/44/4/348.refs.html

>> Version of Record - Sep 28, 2011

OnlineFirst Version of Record - Jun 17, 2011

What is This?
Exploring Dynamic Assessment as a Means of Identifying Children At Risk of Developing Comprehension Difficulties

Amy M. Elleman¹, Donald L. Compton¹, Douglas Fuchs¹, Lynn S. Fuchs¹, and Bobette Bouton¹

Abstract
In this study, the authors explore a newly constructed dynamic assessment (DA) intended to tap inference-making skills that they hypothesize will be predictive of future comprehension performance. The authors administered the test to 100 second-grade children using a dynamic format to consider the concurrent validity of the measure. The dynamic portion of the assessment comprised teaching children to be “reading detectives” by using textual clues to solve what was happening in the story. During the DA children listened to short passages and answered three inferential questions (i.e., one setting, two causal). If children were unable to answer a question, they were reminded what a reading detective would do and given a set of increasingly concrete prompts and clues to orient them to the relevant portion of text until they could answer the question correctly. Results showed that the DA correlated significantly with a standardized measure of reading comprehension and explained a small but significant amount of unique variance in reading comprehension above and beyond vocabulary and word identification skills. In addition, results suggest that DA may be better than the standardized measure of reading comprehension at identifying intraindividual differences in young children’s reading abilities.

Keywords
assessment, comprehension, reading, inference, reading disabilities

Although much of the research in reading disabilities (RD) has focused on problems resulting from poor word identification, there are a substantial number of children who have difficulty understanding what they read despite having adequate word identification skills (e.g., Cain & Oakhill, 2007; Nation & Snowling, 1997; Yuill & Oakhill, 1991). According to Gough and Tunmer’s (1986) simple view of reading, reading is the product of word identification and linguistic comprehension. This framework can be used to classify poor readers into three subtypes: (a) word recognition problems only (i.e., poor decoder or dyslexic), (b) a specific comprehension deficit only (i.e., poor comprehender), and (c) a combination of problems with decoding and comprehension (i.e., garden variety poor reader). It is estimated that poor comprehenders compose 3% to 10% of school-age children (Catts & Compton, 2009; Leach, Scarborough, & Rescorla, 2003). Early identification and treatment of students with poor comprehension has the potential to mitigate later reading problems. Yet these students are often overlooked in the primary years.

One reason for the delay in identification of poor comprehenders may be the nature of reading tasks used in the primary grades. Early reading measures tend to overemphasize word identification skills at the expense of reading comprehension skills, and in doing so comprehension skill assessment is often constrained by decoding (e.g., Catts, Fey, Zhang, & Tomblin, 1999; Francis, Fletcher, Catts, & Tomblin, 2005; Keenan, Betjemann, & Olson, 2008). Reading comprehension assessments constrained by students’ word recognition abilities are unlikely to identify problems that may occur when the text complexity increases to match the students’ linguistic abilities (Catts, Hogan, & Adlof, 2005). Thus, insensitivity of early reading measures to detect comprehension problems may contribute to the delayed identification of students with specific deficits in comprehension.

Researchers have attempted to use listening comprehension as an unconstrained proxy of reading comprehension...
skill for younger and less skilled readers (e.g., Catts, Adlof, & Weismer, 2006). Although the use of listening comprehension measures seems to be a viable alternative for assessing reading comprehension independent of word identification, general listening comprehension measures have not been successful in adequately differentiating students who will later develop reading comprehension problems from those who will not (e.g., Catts et al., 2006; Compton, Fuchs, Fuchs, Elleman, & Gilbert, 2008). One possible reason these tests were ineffective at discriminating late-emerging poor comprehenders may be that they tap general comprehension skills instead of specific reading comprehension skills that become increasingly important in the later grades. Recent research indicates that children’s ability to generate inferences is highly correlated across different modalities (i.e., aural, written, and televised stories), predicts later reading comprehension (e.g., Kendeou, Bohn-Gettler, White, & van den Broek, 2008), and thus may be more appropriate for identifying children who will develop comprehension problems. With this in mind, we developed a listening measure to tap children’s inference generation, a skill considered crucial to later comprehension.

We decided to develop a dynamic assessment (DA) because we believed this type of assessment would allow us to capture and predict students’ responsiveness to comprehension instruction better than a traditional comprehension measure. Scores on traditional assessments may underestimate a child’s ability to benefit from instruction because they reflect what the child has learned prior to testing instead of the child’s potential for learning. When using a DA format, the examiner provides feedback to facilitate the student’s performance. The amount of feedback required for the student to solve a task is a measure of how responsive he or she is to the instruction provided. There are many different kinds of DA models (for a review, see Grigorenko & Sternberg, 1998). For our DA, we decided to use the graduated prompts model developed by Campione and Brown (1987) because this model requires the use of standardized feedback, which has been shown to be more predictive of academic achievement than DA models without such standardization (Caffrey, Fuchs, & Fuchs, 2008). In this model, if a student answers an item incorrectly, the examiner provides the student with a predetermined set of increasingly explicit hints until the student can perform the task independently (Campione & Brown, 1987).

In developing the task for the DA, we sought a task that could (a) predict general comprehension skill, (b) differentiate good and poor comprehenders, (c) eliminate the confound of prior learning (i.e., a task not typically taught to young children), and (d) be learned within a few trials administered in a single testing session. One well-researched skill we felt met our criteria was inference generation. The ability to generate inferences is considered critical to understanding text (e.g., Cain & Oakhill, 2007; Kintsch & Kintsch, 2005; Trabasso & van den Broek, 1985). To gain the full meaning of a text, a reader must first consider the information stated explicitly in the text and then create a coherent mental representation of it (Kintsch, 1998). The creation of this mental model requires the reader to integrate information across the text as well as with his or her prior knowledge. The ability to make inferences has been shown to consistently differentiate good and poor comprehenders (see Cain & Oakhill, 2007; Yuill & Oakhill, 1991). We decided to use a metacognitive-oriented inference intervention that focused on selecting and using textual clues to make inferences because this type of instruction has been shown to be effective on near transfer measures of inferential comprehension within relatively short periods (i.e., less than 4 hr; Reutzel & Hollingsworth, 1988; Winne, Graham, & Prock, 1993; Yuill & Josceleyne, 1988).

In sum, our long-term goal is to use DA to identify children at risk for developing RD because of comprehension problems. In this study, we take the first step in this process by exploring the measure’s reliability and validity. We asked the following questions: (a) What is the internal consistency of the measure? (b) What is the correlation of DA with a validated reading comprehension measure, word identification measures, and vocabulary? and (c) How much unique variance does the dynamic test explain in a validated reading comprehension measure after considering word identification and vocabulary? We also considered how each of the tests classifies students according to the simple view of reading.

**Method**

**Study Design**

The orally administered dynamic test consisted of 7 passages and 21 test items (3 for each passage) with the addition of one training passage. No data were collected for the items pertaining to the training passage. The passages for the DA were presented over five phases (see Figure 1).

**Participants**

We administered the DA to 100 second-grade students across 24 classrooms in nine public schools in Nashville, Tennessee, who were selected from a larger pool of students (N = 391) participating in a longitudinal study. From this larger sample, we selected a representative sample of readers (25 high, 50 average, and 25 low) based on a factor score derived from their first-grade scores on the Test of Word Reading Efficiency (Torgesen, Wagner, & Rashotte, 1997) and the Woodcock Reading Mastery Test–R/NU (WRMT-R/NU; Woodcock, 1998) subtests of...
Word Identification (WID), Word Attack, and Passage Comprehension (PC). The mean age of the sample at testing during second grade was 8 years and 3 months. Of the sample, 55% was female, 53% received free or reduced-price lunch, and 12% received special education services. The racial makeup of the sample was 36% African American, 42% Caucasian, 8% Hispanic, 8% Kurdish, 2% Asian, and 4% Other.

Measure Development

Passages. The test included seven short passages that were read aloud to the children. Five of the stories were created to be equal in difficulty, and two passages were created to assess transfer and avoid ceiling effects. One of the transfer passages was created with a lower level of cohesion because research has shown that lower levels of cohesion negatively affect reading comprehension for poor readers (e.g., McNamara, Kintsch, Songer, & Kintsch, 1996). The low-cohesion passage contained a higher number of pronouns, fewer causal connectives, and more filler text between inferences. The other transfer passage was created to assess students’ inference generation in expository text. This text was adapted from a third grade text about the rainforest (Myers, 1999). Early in test development, we considered using expository text to simulate the textual changes that occur in late elementary school. However, we decided that the difficulty and unfamiliarity of expository tests would likely make the task too difficult for second graders and preclude us from assessing any differences resulting from learning, especially within a single session. Therefore, except for the one transfer passage, all other passages were narrative.

The narrative passages were based on passages developed by Yuill and Joscelyne (1988). The passages were crafted so that the setting of the story and pieces of information vital to understanding the story were never explicitly stated. These omissions required the children to make an inference about the story’s setting and causal inferences for what was happening in the story. With consideration of the abilities of the second-grade children in our sample, we decided to use causal inference, not only because of their prominent role in understanding narrative text but also because they have also been shown to be easier than other types of inferences (e.g., Bowyer-Crane & Snowling, 2005; Trabasso & van den Broek, 1985).

The passages contained 160 to 217 words and ranged from a third-to fifth-grade level as indexed by the Flesch–Kincaid readability formula. Passages in the test, except for the transfer phase, were equated on vocabulary and syntax using the Coh-Metrix indices developed by Graesser, McNamara, Louwerse, and Cai (2004). The passages were created to be similar in content and complexity as the texts used in late elementary school. The listening format of the test allowed us to use passages more aligned with the students’ oral language abilities rather than their word identification skills. We also attempted to control for the impact of prior knowledge because differences in prior knowledge have been shown to affect inference generation (e.g., Hansen & Pearson, 1983). We, therefore, created stories with familiar topics (e.g., climbing toddlers, messy friends, mean bullies) and settings (e.g., grocery store, classroom, park).

Items. Three open-ended questions were created for each passage for a total of 21 items. For the first question after each story, students were required to make a setting inference. The setting questions were considered fairly easy and were included to avoid floor effects for students with poorer inference skills. For the other two questions, we asked students to make causal inferences. Research suggests that readers have more difficulty making inferences when the information needed to be integrated to make the inference is distally rather than proximally located in the text (e.g., Ackerman, Jackson, & Sherill, 1991). These difficulties are more pronounced for poor comprehenders (e.g., Ehrlich, Remond, & Tardieu, 1999). Therefore, to create a range of difficulty in the items and to better differentiate good and poor comprehenders, we developed two causal inference questions for each passage. The first causal inference question required students to integrate clues in the text across shorter amounts of text and the other required integration across longer amounts of text.
Inference instruction. After pretest, students received instruction to improve their inference skills. Instruction was designed to increase students’ inference skills by teaching them to find and use important information in the text (Reutzel & Hollingsworth, 1988; Winne et al., 1993; Yuill & Joscelyne, 1988). During the inference instruction phase of the test, students were taught to be “reading detectives” by identifying clues in the text to help them figure out what is happening in the stories. After discussing the similarities between good readers and detectives, the examiner explained that good reading detectives pay attention to repeated information, use clues across all parts of the text, and keep looking for clues until the story makes sense. After this instruction, the examiner read an instructional passage aloud and modeled how to use the clues to solve what was happening in a story. The examiner demonstrated how to use the clues in the story to answer three inference questions similar to those used for the other passages.

Prompts. Prompts were created for each of the nine items in the dynamic phase of the test. The majority of prompts consisted of reminding the student how to be a reading detective and orienting him or her to clues in the story. We also added a prompt that consisted of rereading the story. Even though the story was present for the children to refer to, we wanted to make sure that children with poor word identification could make use of the text to help them remember events or details of the story. For each item the children could not answer, they were provided with a prompt in the form of a highlighted text clue that the examiner pointed to and read aloud. The clues were presented from least to most helpful for making the inference. The last prompt in each series of prompts consisted of a summary of all of the clues presented in the story. Students were presented each prompt until they answered the question correctly or the prompts were exhausted. An example of a passage and the prompting procedure is provided in Figure 2.

Measures

Vocabulary (VOC). VOC was measured using the Vocabulary subtest of the Wechsler Abbreviated Scale of Intelligence (WASI; Psychological Corporation, 1999). The WASI is a validated, norm-referenced test for ages 6 to 89 years. This subtest contains 42 items that require the student to name pictures for the first 4 items and then define words that are visually and orally presented. The internal consistency for the Vocabulary subtest exceeded .90 and the test–retest reliability exceeded .86 for the children’s sample.

WID. The WID subtest of the WRMT-R/NU (Woodcock, 1998), a norm-referenced test, was used to assess word identification skills. Children read a list of increasingly difficult words. Split-half reliability for WID exceeded .94 for the second-grade sample.

Woodcock PC. Reading comprehension was assessed using the PC subtest of the WRMT-R/NU (Woodcock, 1998). This subtest uses a modified cloze format to assess reading comprehension. The child silently reads a short passage and identifies the missing word in the passage. The split-half reliability for the second-grade sample exceeded .90.

Procedure

Administration. Students were individually assessed over two sessions within 2 weeks in early May. The DA was given in one session of 25 min to 1 hr depending on the ability of the child. All of the remaining measures were given in another session that lasted up to 1 hr. At the start of the DA, the examiner explained that he or she was going to read some tricky stories that do not tell everything that is happening in them. The students were told that each of the stories contained clues to help them figure out what is happening. Students were also told that they could have any part of the story or questions reread to them. Throughout the test, the passages were available to the student to refer to when listening to the story or answering the questions. The student was encouraged to follow along in the text while listening to the story. Next, the examiner presented the pretest passage and items to the student. After the pretest, the examiner presented the lesson on inference generation (i.e., reading detective lesson on how to find and use clues in a story) using a practice story. Then, the examiner led the student through the dynamic phase of the test that included nine items over three passages. If the student answered an item incorrectly, the examiner provided prompts until the student answered correctly or the prompts were exhausted. Last, the examiner presented the posttest story and two transfer stories. The examiner provided no prompts for items on these stories.

Scoring. We used three scores to capture different aspects of student performance: (a) the number of prompts needed during the 9 items in the dynamic phase of the test, which could range from 0 to 45 (15 possible for each passage); (b) a transfer score reflecting the number of correct responses to the 6 items on the low-cohesion and expository texts, which could range from 0 to 30; and (c) a total score based on all seven stories, which could range from 0 to 105. To create a total score, items with prompts were rescoring (see Table 1). Static items (i.e., no prompts provided) answered incorrectly were scored as 0. If answered correctly, the static items were scored 4 for setting questions, 6 for near-causal questions, and 5 for far-causal questions, thereby weighting correct unprompted answers equivalently across items in the dynamic and static phases of the test.

Fidelity and reliability of administration. The DA was administered by 11 graduate students trained in the administration of educational assessments. Examiners received training in administration and scoring for 3 hr and were required to
Jenny was a very active toddler. She climbed on everything at home. Last week Jenny used the drawers in the kitchen to climb up on the counter, because she wanted to get a cookie shaped like a tiger. Jenny loved tigers. Jenny had an older brother named Tyrone. Today, Jenny was going to the store with her mother and Tyrone. Jenny hated to ride in the shopping cart, so Tyrone asked if he could take her to look at the toys in the cereal aisle. Their mother warned Tyrone to hold Jenny’s hand, so Jenny wouldn’t get into anything. As Jenny and Tyrone walked past the cereal boxes, Jenny pointed up at the top shelf to a box with a tiger on it and clapped. Tyrone took Jenny over to the toys. Jenny wasn’t interested in the toys, so she pulled her hand away from Tyrone. She ran down the long aisle. All of a sudden, Tyrone heard some crashing sounds. Jenny was crying.

Sample Questions and Prompts

<table>
<thead>
<tr>
<th>Question 1 (setting)</th>
<th>Question 2 (causal)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>1. Where are Jenny and Tyrone at the end of the story?</strong></td>
<td><strong>2. What made the crashing sounds?</strong></td>
</tr>
</tbody>
</table>

**Prompt #1:** “Let’s be reading detectives and use the clues to help us figure out where they are. Here the story says, “she ran down the long aisle.”

**Prompt #2:** Here is another clue to help you figure out where Jenny and Tyrone are. The story says, “Jenny hated the shopping cart.”

**Prompt #3:** Here are some more clues. The story says “cereal aisle” and it says “cereal boxes.”

**Prompt #1:** “The story doesn’t really tell you what made the crashing sounds. Sometimes when I can’t figure out what’s going on in a story, I reread it and look for clues that might help. I will reread the story. Be a reading detective and look for clue words or sentences that might help you figure out what made the crashing sounds.

**Prompt #2:** Here are some clues to help you figure out what made the crashing sounds. The story says, “Tyrone took Jenny over to the toys. Jenny wasn’t interested in the toys, so she pulled her hand away from Tyrone.” And here it says, “Jenny was crying.”

**Prompt #3:** Here is another clue. Remember reading detectives have to think really hard about the clues. The story says, “Their mother warned Tyrone to hold Jenny’s hand, so Jenny would not get into anything.” It also says, “She pulled her hand away from Tyrone. She ran down the long aisle.” And here it says, “Tyrone heard some crashing sounds.”

**Prompt #4:** Here are some more clues. The story says, “Tyrone asked if he could take Jenny to look at the toys in the cereal aisle,” and it says, “As Jenny and Tyrone walked past the cereal boxes, Jenny pointed up to a box with a tiger on it and clapped.” We can be reading detectives by looking for clues earlier in the story. Earlier in the story it says, “Jenny loved tigers.” Remember reading detectives put all of the clues together to figure out what’s going on.

**Prompt #5:** A good reading detective remembers all of the clues and puts them together to make the story make sense. Let’s go over the clues we have so far about what made the crashing sounds. We know that Jenny ran away from Tyrone, because she wasn’t interested in the toys. We know their mother warned Tyrone to hold Jenny’s hand, so she wouldn’t get into anything. We also know that Jenny loved tigers and clapped when she saw a cereal box with a tiger on it. And we know that Jenny was crying.

**Figure 2.** Example of a passage and the prompting procedure

demonstrate competency by administering the test in a mock session with 95% fidelity for administration and scoring. If examiners did not reach the criterion, they were retrained and retested until they met criterion. All of the directions and dialogue were scripted to ensure standardization across examiners. An answer key was provided to each examiner,
so he or she could determine if a child had answered a question correctly or required another prompt. Sometimes children answered questions ambiguously. Answers children might provide were included in the administration guide. When an answer was not provided and the examiner felt that the answer was close, the examiner was instructed to ask, “Can you tell me more about . . .?”. All answers and any extra dialogue between the examiner and student were transcribed on the scoring sheet. In addition, all sessions were tape-recorded. Administration fidelity was assessed by completing a component checklist. Administration of each story was evaluated based on whether the examiner correctly presented the questions, prompts, and queries as well as how he or she read each story. Administration and scoring reliability was calculated using percentage agreement (i.e., percentage agreement = agreements / agreements + disagreements). Overall reliability of administration and scoring for the DA was 97.8%. Scoring and reliability were calculated for 20% of the other tests in the battery and ranged from 89.1% to 96.0%. All errors for each test were reconciled by one of the authors.

Data Analysis

First, reliability was evaluated by determining the internal consistency of the initial items in the measure using Cronbach’s alpha. Next, we considered the concurrent validity of the dynamic measure. We compared outcomes on the DA (i.e., number of prompts, transfer score, and total score) to the PC subtest and other measures related to reading comprehension including the VOC and the WID. Next, we explored how much variance the DA could explain in the PC above and beyond what explained by word identification and vocabulary. Finally, because the DA was developed to fill a gap not being addressed by current measures of reading comprehension, we decided to explore differences between DA and PC for classifying students according to the Gough and Tunmer’s (1986) simple view of reading. To do this, we converted each student’s PC, DA, and WID raw scores to z scores. Next, we plotted each student’s score on WID against scores on his or her score on the PC and then against scores on the DA.

Results

The internal consistency of DA as measured by Cronbach’s alpha was .76, indicating sufficient reliability for the test. To consider the concurrent validity of the DA, we compared it to PC, a validated measure of general reading comprehension, and VOC and WID, measures tapping skills important to reading comprehension. Descriptive information and correlations are provided in Table 1. It should be noted that one student did not receive the second battery of tests, so all results including measures other than the DA were based on a sample of 99 children. Correlations among the variables showed that each of the measures correlated significantly with one another. As expected, the number of prompts a student required on the DA was negatively related to the student’s general comprehension but not as strongly associated with PC as the total score. We, therefore, used the total score instead of prompts or transfer score in all of the subsequent analyses.

As indicated by comprehension research using similar measures (Keenan et al., 2008), we found that the PC and WID subtests were highly correlated. We also found a strong relationship between the DA and WID. We were surprised by this because we had tried to control for word identification by administering the DA as a listening measure. We believed that this relationship might be mediated by vocabulary, so we conducted a regression analysis with the DA as the dependent variable. We found that WID continued to correlate significantly with the DA even after controlling for VOC ($p = .23$, $p < .05$).

Next, we turned to exploring the unique relationship between DA and PC. First, we conducted a regression analysis to determine how much variance the DA accounted for in PC after considering the variance explained by word identification and vocabulary (see Table 3). WID and VOC were entered first (Model 1), with DA entered next (Model 2). The total amount of variance explained increased from 74% to 78% across models indicating that the DA uniquely explained 4% unique variance in PC. Note that after entering

Table 1. Scaled Scores on Items With Prompts for Total Score

<table>
<thead>
<tr>
<th>Points</th>
<th>Correct initial answer</th>
<th>1 prompt required</th>
<th>2 prompts required</th>
<th>3 prompts required</th>
<th>4 prompts required</th>
<th>5 prompts required</th>
<th>6 prompts required</th>
<th>7 prompts required</th>
<th>8 prompts required</th>
<th>9 prompts required</th>
<th>10 prompts required</th>
<th>11 prompts required</th>
</tr>
</thead>
<tbody>
<tr>
<td>Setting question</td>
<td>Correct initial answer</td>
<td>4</td>
<td>3</td>
<td>2</td>
<td>1</td>
<td>0</td>
<td>Correct initial answer</td>
<td>6</td>
<td>5</td>
<td>4</td>
<td>3</td>
<td>2</td>
</tr>
<tr>
<td>Causal—near question</td>
<td>Correct initial answer</td>
<td>7</td>
<td>6</td>
<td>5</td>
<td>4</td>
<td>3</td>
<td>2</td>
<td>1</td>
<td>Correct initial answer</td>
<td>9</td>
<td>8</td>
<td>7</td>
</tr>
</tbody>
</table>

Downloaded from dx.doi.org at MIDDLE TENNESSEE STATE UNIV on July 18, 2012
Table 2. Correlations and Descriptive Statistics for the Reading Measures, Vocabulary, and Dynamic Assessment (DA)

<table>
<thead>
<tr>
<th></th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Passage comprehension</td>
<td>—</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2. Word identification</td>
<td>.84</td>
<td>—</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3. Vocabulary</td>
<td>.67</td>
<td>.63</td>
<td>—</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4. DA prompts</td>
<td>-.59</td>
<td>-.46</td>
<td>-.63</td>
<td>—</td>
<td></td>
<td></td>
</tr>
<tr>
<td>5. DA transfer</td>
<td>.50</td>
<td>.44</td>
<td>.47</td>
<td>-.47</td>
<td>—</td>
<td></td>
</tr>
<tr>
<td>6. Total DA</td>
<td>.70</td>
<td>.58</td>
<td>.70</td>
<td>-.85</td>
<td>.72</td>
<td>—</td>
</tr>
</tbody>
</table>

M     23.04  55.19  27.54  14.27  8.62  51.01  
SD    5.49   13.61  6.99   9.19   6.93   18.02  
Standard score 95 98 100 — — —  
Range 9–35 12–90 5–46 1–36 0–30 9–90  
Max    68   106   56   45   30   105  

Note: N = 99. All analyses were conducted with raw scores. All correlations are significant, p < .01.

Table 3. Hierarchical Regression Analysis Estimating the Unique Variance Associated With the Dynamic Assessment (DA) Using the Woodcock Reading Mastery Test–R/NU Comprehension Subtest as the Dependent Measure and Controlling for Word Identification and Vocabulary

<table>
<thead>
<tr>
<th></th>
<th>B</th>
<th>β</th>
<th>t</th>
<th>p</th>
<th>Adj. R² of Model</th>
</tr>
</thead>
<tbody>
<tr>
<td>Model 1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Constant</td>
<td>2.50</td>
<td>1.93</td>
<td>.05</td>
<td>.74</td>
<td></td>
</tr>
<tr>
<td>Word identification</td>
<td>0.28</td>
<td>.69</td>
<td>10.36</td>
<td>.001</td>
<td></td>
</tr>
<tr>
<td>Vocabulary</td>
<td>0.19</td>
<td>.24</td>
<td>3.56</td>
<td>.01</td>
<td></td>
</tr>
<tr>
<td>Model 2</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Constant</td>
<td>2.96</td>
<td>2.43</td>
<td>.02</td>
<td>.78*</td>
<td></td>
</tr>
<tr>
<td>Word identification</td>
<td>0.26</td>
<td>.63</td>
<td>9.73</td>
<td>.001</td>
<td></td>
</tr>
<tr>
<td>Vocabulary</td>
<td>0.07</td>
<td>.09</td>
<td>1.17</td>
<td>.24</td>
<td></td>
</tr>
<tr>
<td>DA total</td>
<td>0.082</td>
<td>.022</td>
<td>3.81</td>
<td>.001</td>
<td></td>
</tr>
</tbody>
</table>

Note: N = 99. All analyses conducted with raw scores. *Significant ΔF(1, 95) = 14.52, p < .001.

Figure 3. Scatterplots of dynamic assessment (DA) by word identification (WID) and passage comprehension (PC) by WID.

PC than the relationship between WID and the DA, indicating that more children have substantial intraindividual differences in their reading skills on the DA than on the PC. Although we cannot rule out the possibility that the spread in scores seen with the DA may be the result of measurement error, the pattern of correlations gives some credence to the idea that the differences displayed may be the result of the differences in the constructs underlying the tests.

To better illustrate these classification differences and consider the subgroup we were most interested in, poor comprehenders, we cut the distribution on each measure to correspond to low (z score ≤ –1) or average (z score ≥ –1). (Note that although a cutoff score of –1 is commonly used to identify poor readers, this score is arbitrary and these groupings would change if the cutoff score were moved.) Using this system, 10 children were identified as having poor DA skill with adequate WID skill (i.e., poor...
comprehender subtype), whereas only 3 children were identified as having low PC skill with adequate WID skill.

Discussion

Providing early intervention for children with poor comprehension is dependent on accurate identification. Recently, researchers have turned a critical eye toward standardized measures of reading comprehension, asking important questions about what these tests are actually measuring. There is a concern that the insensitivity of reading comprehension measures at the primary level may be impeding early identification and intervention of reading comprehension deficits. Addressing some of the concerns, the RAND Reading Study Group (2002) suggested guidelines for developing measures for the identification of poor comprehenders including that comprehension tests should be driven by reading theory, reliable and valid at the item level, sensitive to developmental shifts in reading, and informative to practitioners. With this in mind, we designed a DA to help identify children at risk for developing RD because of comprehension difficulties.

Results of the reliability and concurrent validity of the DA were encouraging. The internal consistency of the initial items indicated that the reliability for the DA was adequate. In addition, we found the DA total score correlated highly with PC, suggesting that the DA is measuring a similar construct of comprehension as PC. As expected, the number of prompts a student required on the DA was negatively related to scores on the PC. This relationship was not as strong as the relationship between the total score on the DA and the PC. This may be because the total score provided more information than the subset of items that contained prompts.

We developed the DA as a listening measure in an attempt to isolate comprehension from word identification skill, so we were surprised by the strong relationship between DA and WID. We were prompted to check if vocabulary was mediating this relationship because other researchers have found a substantial amount of shared variance between word identification and language in studies with young children (e.g., Keenan et al., 2008). We, therefore, checked to see if vocabulary was mediating the relationship between the DA and WID. Although most of the variance between the DA and WID was shared with VOC, WID remained significantly correlated with DA after controlling for vocabulary. It is unclear what could be influencing this relationship, but one factor may be that students in the study had the text available to them at all times. Students who were better readers may have benefited from this presentation by taking advantage of the opportunity to look back through the text to answer the questions. Poorer readers may have had more difficulty using the text in this way or may have been more inclined to only listen to the stories.

We also found that DA explained 4% unique variance above and beyond vocabulary and word identification on the PC. Although the unique variance explained by the DA does not seem very substantial, it does indicate that the DA is picking up something that word identification and vocabulary are not addressing in the PC. With PC and WID being very strongly correlated, it is interesting that the DA picks up any unique variance beyond word identification and vocabulary. We created the DA to fill a gap not being addressed by traditional reading comprehension measures for primary students. Many of these measures have been found to be dependent on word identification, not comprehension. The constructs the DA was created to capture (i.e., responsiveness to instruction, inferential comprehension, and listening comprehension) are different from the constructs underlying many current reading comprehension measures. Therefore, many of the constructs intended to be captured by the DA are not represented in the unique variance because the PC does not address these constructs.

To explore how these two tests may be different, we examined differences between PC and DA in classifying student according to the simple view. In this comparison, the DA showed more differentiation among word identification and comprehension skills than the PC. This finding is consistent with the research discussed earlier that showed that the PC relies heavily on decoding skills, especially for younger kids. These results suggest that the DA may be better than PC at identifying intraindividual differences in young children’s reading abilities. Although our classification example was exploratory, the pattern of results was interesting. A comparison of the DA and WID identified more students exhibiting a poor comprehender profile than the PC. It is also interesting to note that many children across reading levels show marked differences in their word identification and comprehension abilities. We believe the DA may be capturing skills that other comprehension tests do not address and therefore may be better suited to finding students likely to have reading problems because of comprehension deficits. Ultimately, this hypothesis can be tested only by addressing the predictive validity of the DA. To accomplish this, we plan to retest the students in fourth grade.

Limitations

The design of this study limits the conclusions that can be drawn about the importance of the dynamic aspect of the DA. The first concern is that an evaluation of the effectiveness of the inference instruction and feedback was not conducted. The amount of time and resources required for administering the test needs to be considered. There
is a balance that must be maintained between the extra information gained from the dynamic test and the resources required to administer it. The inference training required time that may not be necessary for the test to predict comprehension deficits. In addition, the administration of the prompts requires the test to be administered to each child individually. The design of this study did not allow us to evaluate the effectiveness of the prompts or their proper ordering. Although prior intervention studies have found that instruction orienting children to relevant information increases reading comprehension, we have no way to know if providing clues helped the children make the inferences. It is possible that the pattern of responses could be an artifact of allowing multiple opportunities to answer a question.

In addition, the design of this study did not allow us to adequately assess the relative contributions of various aspects of the DA. For example, although the PC and DA are correlated, they classify students differently. Are the differences found between the DA and PC because DA is tapping inferential comprehension, listening comprehension, responsiveness to learning, or a combination of some, or all, of these aspects? In a previous study, we found that although the listening comprehension variable looked promising for predicting students with late-emerging RD, it produced too many false positives (Compton et al., 2008). Thus, it is likely that the DA will have to explain variance above and beyond that attributed to the listening method to help in the prediction of late-emerging poor comprehenders. The inclusion of a listening comprehension measure and validated measure of inferential comprehension in our test battery would have been beneficial for teasing apart effects because of method and test content. The use of the PC subtest was not ideal, and further research with the DA will have to be conducted with other comprehension measures.

**Conclusion**

The need for early identification and intervention for poor comprehenders is underscored by recent findings that comprehension abilities exist across different media (Kendeou et al., 2008). These findings suggest that the problems exhibited by poor comprehenders could be pervasive and extend beyond the written word. If this is so, poor comprehenders are likely to have difficulties in many areas of their lives, in and outside of school. Constructing reliable and valid tests for the early identification of these children will require a new consideration of how tests should be constructed and what they should measure. We hope that the DA, used in combination with other tests, will be helpful in differentiating young children who are at risk of developing comprehension difficulties. In this first attempt in examining the reliability and validity of the measure, we found some evidence for the internal reliability and construct validity of the DA. Although promising, more work will need to be conducted to determine the measure’s predictive power, to isolate and adequately capture children’s responsiveness to the instruction provided, and to determine the measure’s relative utility among other tests of early comprehension before any definitive recommendations can be made regarding its use.

**Declaration of Conflicting Interests**

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

**Funding**

This research was supported in part by Grant R324G060036 from the U.S. Department of Education, Institute of Education Sciences and Core Grant HD15052 from the National Institute of Child Health and Human Development to Vanderbilt University. Statements do not reflect the position or policy of these agencies, and no official endorsement by them should be inferred.

**References**


About the Authors

Amy M. Elleman, PhD, is a research associate in special education at Peabody College of Vanderbilt University. She is interested in the early identification and remediation of reading comprehension difficulties.

Donald L. Compton, Ph D, is an associate professor in special education at Peabody College of Vanderbilt University. His research focuses on modeling individual differences in reading acquisition.

Douglas Fuchs, PhD, Nicholas Hobbs Professor of Special Education and Human Development at Peabody College of Vanderbilt University, studies reading instruction and assessment, with a focus on learning disabilities.

Lynn S. Fuchs, PhD, Nicholas Hobbs Professor of Special Education and Human Development at Peabody College of Vanderbilt University, studies mathematics and reading instruction and assessment, with a focus on learning disabilities.

Bobette Bouton, MEd, is a doctoral student at the University of Georgia.