Computer Based Assessment System-Reading (CBAS-R):

Technical Report on Field Test Results 2009-2010

Allison McCarthy
Amanda Pike
Theodore Christ
University of Minnesota

Date Initiated: 2/2010
Finalized Draft: 11/23/10
Reviewed by Advisory Board: TBA

Preparation of this technical report was supported in part by a grant from the Office of Special Education Programs, U.S. Department of Education (H327A090064). Opinions expressed herein do not necessarily reflect the position of the U.S. Department of Education.
Abstract

The sampling plan and data collection to establish the item parameters for CBAS-2 items are presented. Data were collected in the Fall and Winter of 2009-2010 at seven different schools in Minnesota. 638 items from the CBAS-2 item bank were administered; 16 items were selected and used as linking items, while 622 additional items were parameterized and retained for the subsequent version of CBAS-R. This report summarizes details of data collection including demographics of samples, identification of linking items, test item selection, and rationale for stratified sampling procedures.

Keywords: sampling, CBAS-2, plan, 2009-2010
## Research Log

<table>
<thead>
<tr>
<th>Activity</th>
<th>Lead Person</th>
<th>Timeframe</th>
</tr>
</thead>
<tbody>
<tr>
<td>Recommendations to PI’s</td>
<td>Allison McCarthy</td>
<td>2/10</td>
</tr>
<tr>
<td>Writing</td>
<td>Allison McCarthy, Amanda Pike</td>
<td>2/10-11/10</td>
</tr>
<tr>
<td>Review</td>
<td>Theodore Christ</td>
<td>11/2/10</td>
</tr>
</tbody>
</table>
Table of Contents

Abstract .................................................................................................................................................. 2
Introduction ............................................................................................................................................ 5
  Purpose of CBAS-R ............................................................................................................................... 5
  Field Testing CBAS-R .......................................................................................................................... 5
Method .................................................................................................................................................... 6
  Participants ........................................................................................................................................... 6
Initial Field Testing to Establish Linking Items ...................................................................................... 7
Field Testing to Establish Item Parameters ............................................................................................ 9
Procedure: Data Collection Overview ................................................................................................. 10
Procedure: Sampling Plan .................................................................................................................... 10
  Fall 2009 Data Collections .................................................................................................................. 10
    October (Schools A & B) .................................................................................................................... 10
    November (Schools C & D) ................................................................................................................. 11
    December (School E) ......................................................................................................................... 11
Winter 2010 Data Collections ................................................................................................................ 12
  January (School A) .............................................................................................................................. 12
  February (Schools C, D, & G) .............................................................................................................. 12
  March (Schools B, E, & F) .................................................................................................................... 13
Results .................................................................................................................................................... 13
References .............................................................................................................................................. 14
Appendix ................................................................................................................................................ 15
CBAS-R: Technical Report on Field Test Results 2009-2010

Purpose of CBAS-R

The Computer-Based Assessment System for Reading (CBAS-R) is a computer administered assessment of reading achievement for use in kindergarten through fifth grade. CBAS-R outcomes will enable continuous evaluation of the level and rate of reading achievement within and across the primary grades. The application of contemporary psychometric theory (3-Parameter Item Response Theory) has been used to support the development of CBAS-R as a computer adaptive test (CAT). The advent of CAT-based assessments has the potential to confer substantial benefits to teachers and educators. The CBAS-R targets reading achievement across five critical sub-domains put forth by the National Reading Panel (2000): concepts of print, phonemic awareness, phonics, vocabulary, and comprehension. It also provides sufficient flexibility to assess within the sub-domains and derive profiles of achievement across domains – and outcomes will be equated both horizontally and vertically (i.e., within and between grades).

Field Testing CBAS-R

The CBAS-R team sought to field-test the new CBAS2 item bank (Fall 2009) in hopes of having a better and more representative bank of items than the previous (CBAS) bank of items. Data were collected in the Fall and Winter of 2009-2010 at seven different schools in Minnesota, with similar methods used in the 2007 field testing data collection (See Ayodele, Jiban, Christ, & McCarthy: Tech. Rep. No. 2).
Method

Participants

Students in kindergarten through fifth grade were drawn from seven schools in the suburbs near Minneapolis, Minnesota. As the goal was to test a large number of students (300 per item), CBAS personnel wanted to ensure that the majority of students in schools would be able to participate without getting frustrated. Thus, English Language Learner (ELL) students would not be the best population to participate as the tests may have caused frustration due to the large amount of reading required. School demographics are presented in Table 1. Sample sizes by school and grade are presented in Table 2.

Schools were identified for participation based on a high rate of non-ELL students. Principals were emailed inquiring about their interest to participate. The Project Coordinator corresponded with principals to gauge their interest in participation. In some cases, principals recommended that CBAS staff contact individuals at the administration level (e.g., Superintendent) to confirm that it was appropriate for their school to participate. Once confirmation was received from administration staff or the principal, CBAS personnel proceeded with scheduling the data collections.

Initial Field Testing to Establish Linking Items

Item Response Theory (IRT) requires that each field tested item be sampled on 300 individuals (Weiss personal communication, 2009). Therefore, our goal was to reach 300 students tested per item; some items were tested on more than 300 students due to available participants in schools. Sampling at this rate ensures that all three parameters (a, b, and c) are stable estimates of the item parameters.
The goal of the first data collection was to identify and establish parameters for linking items. Linking items \((N = 16)\) are those items that were included on each subsequent field test as necessary to use a mixed group common item linking procedure (Kolan & Brennan, 2004). Without linking items, examinee responses across alternate items and examinees samples could not be compared and used to estimate item parameters. Linking items were developed and selected to represent all domains as well as the range of item difficulties, which generally correspond with reading performance from kindergarten through fifth grade. Identification of linking items from the first data collection was extremely important because they appear on each subsequent field test. CBAS personnel identified 16 linking items that spanned across the difficulty level and were content-balanced (all linking items can be found in the appendix of this report).

CBAS personnel systematically chose 55 items out of the entire bank of items \((N = 638)\). First, they individually chose what they determined were the best 55 items based on findings from previous administrations during CBAS-R-1 and on literacy research. Any item that both members chose was included in the 55 with the assumption that they were chosen because they were good items. The number of items chosen per domain was calculated based on our a priori model. For example, Concepts of Print and Phonological Awareness develop in kindergarten and first grade. Therefore, the proportion of these item types should be much smaller than the proportion of Comprehension items on tests; Comprehension develops across five grades, not just two. The number of items per grade level was similarly determined.

CBAS staff decided that some items would be too difficult for young students at the first data collection in which 55 items were to be tested. Thus, the administration was structured to have five tests with branching and termination criteria. Each test was content-balanced and
successive tests were sequenced so they became progressively more difficult (item difficulty was based on expert judgment of project personnel). For instance, the first test included 15 items that were considered to be the easiest. Based on performance within each test, students’ administration was either terminated (due to the number of incorrect responses) after a test was completed and prior to subsequent tests, or continued to the next, progressively difficult, test (if enough correct responses were given). If an examinee’s test terminated, then it was assumed that the examinee would not respond to subsequent items at a rate greater than chance.

The results of this initial field testing were used to guide the selection of linking items. These results were not included in the supermatrix or to estimate the final item parameters.

**Field Testing to Establish Item Parameters**

Researchers did not want any students to become frustrated by items that were too difficult. Based on best estimates of the project team, items were sequenced by difficulty and divided into 6 sets of 3 progressively more difficult tests (ex: for Schools A & B there was a Fall set of 3 tests (K-1, 2-3, 4-5) and a Spring set of 3 tests (K-1, 2-3, 4-5)). While all reading domains were assessed at all grade levels, grade level tests were constructed with the particular domain emphasis for that grade in mind. Specifically, personnel believed that Kindergarten and first grade examinees should be assessed in the domains of Concepts of Print, Phonological Awareness, and some Phonics skills. Thus, it seemed reasonable to combine these grades and test them using the same set of items. Similarly, it was expected that examinees in grades two and three and grades four and five were similar in reading development. The most relevant domains for grades two and three seemed to be Phonics and Vocabulary. Finally, the focus for grades four and five was Vocabulary and Comprehension. It is important to note that each grade
was tested on Vocabulary and Comprehension, but at varying levels of difficulty corresponding with grade level.

Excluding the first data collection at School A, both the Kindergarten-First Grade (K-1) tests and the Second-Third Grade (2-3) tests had 39 new items on each. Thirty-one new items were on each Fourth-Fifth Grade (4-5) test. The number of items on the 4-5 tests was reduced due to the majority of items belonging to the Comprehension domain. In general, Comprehension items require more time because more reading is involved. This means that the K-1 and 2-3 tests each had a total of 55 items, including the linking items, and the 4-5 test had a total of 47 items, including the linking items.

**Procedure: Data Collection Overview**

Data were collected using a mobile computer lab ($N = 28$ laptops each). At the time of administration, data collectors directed students to the appropriate computer, located in a test carrel, and provided a brief set of directions. A picture of an individual student administration is presented in Figure 1 of the Appendix. Visual stimuli are presented on a computer monitor and auditory stimuli are presented with padded earphones (padding helps isolate students from extraneous noises and helps ensure that the assessment could be conducted within a populated area). The students receive oral directions from a proctor or teacher followed by automated directions and on-screen demonstrations to guide them through the test. Students responded to CBAS items in sessions of 15 – 30 minutes. Upon completion, students were free to return to their classroom.

**Procedure: Sampling Plan**

**Fall 2009 Data Collections**

October.
School A. The purpose of this initial data collection was to identify linking items (procedure described above). Three hundred sixty four students participated in this data collection. Once linking items were identified ($N = 16$), they were included in all subsequent tests for which items were field tested (with one exception; School B Fall Data collection, 4-5 test).

School B. As another early data collection in which CBAS personnel had no data to aid us in choosing items, items considered to be high quality and representative of a range of difficulty and content for this data collection were selected. Five linking items were unintentionally left out of the 4-5 test. As a result, items from the 4-5 test were added to the Supermatrix once all data collections were completed. The researchers adjusted the parameters appropriately. Five hundred and sixty total students participated in this data collection at this school; specifically, 153 students were administered the K-1 test, 211 students were administered the 2-3 test, and 196 students were administered the 4-5 test.

November.

School C. This elementary school served as a make-up school for items tested at School B. In other words, the goal of 300 students per item was not achieved at School B due to the student population. Students at School C took the same test set as at School B in order to obtain the goal of 300 students per item. Three hundred and sixty two total students participated in this data collection at this school; specifically, 128 students were administered the K-1 test, 123 students were administered the 2-3 test, and 111 students were administered the 4-5 test.

School D. Items were chosen with consideration for difficulty level and content balance. Seven hundred and forty one total students participated in this data collection at this
school; specifically, 243 students were administered the K-1 test, 255 students were administered the 2-3 test, and 243 students were administered the 4-5 test.

**December.**

*School E.* Items were chosen with consideration for difficulty level and content balance. Seven hundred and twenty seven total students participated in this data collection at this school; specifically, 217 students were administered the K-1 test; 238 students were administered the 2-3 test; 272 students were administered the 4-5 test.

**Winter 2010 Data Collections**

**January.**

*School A.* Based on initial item analysis of the October data collection, more items that were very easy (< -2.0) and very difficult items (> +2.0) were needed to expand the range of linking items. As a result, CBAS personnel focused on adding a majority of extremely easy items (in the domains of Concepts of Print and Phonological Awareness) to the K-1 test. The 4-5th grade test included high level Vocabulary items and Comprehension items considered very difficult due to length of passage, vocabulary, and question type. The tests at this school consisted of an entirely new item set. Three hundred and eighty total students participated in this data collection at this school; specifically, 100 students were administered the K-1 test, 130 students were administered the 2-3 test, and 150 students were administered the 4-5 test.

**February.**
School C. Again, this school served as the make-up school for previous fall data collections (i.e., School B, School D, School E). Three hundred and sixty two total students participated in this data collection at this school; specifically, 126 students were administered the K-1 test, 122 students were administered the 2-3 test, and 114 students were administered the 4-5 test.

School D. Items were chosen with consideration for difficulty level and content balance. Tests administered consisted of an entirely new item set. Seven hundred and ten total students participated in this data collection at this school; specifically, 234 students were administered the K-1 test, 252 students were administered the 2-3 test, and 224 students were administered the 4-5 test.

School G. Items were chosen with consideration for difficulty level and content balance. Five hundred and seventy one total students participated in this data collection at this school; specifically, 215 were administered the K-1 test, 174 students were administered the 2-3 test, and 182 students were administered the 4-5 test.

March.

School B. Items were chosen with consideration for difficulty level and content balance. Tests administered at School B were the same as those administered during the winter data collection at School F. Six hundred and fifty three total students participated in this data collection at this school; specifically, 203 students were administered the K-1 test, 238 students were administered the 2-3 test, and 212 students were administered the 4-5 test.

School E. School E served as the make-up school for all data collections in need of more administrations for items. Seven hundred and sixty eight total students participated in
this data collection at this school; specifically, 278 students were administered the K-1 test, 258 students were administered the 2-3 test, and 232 students were administered the 4-5 test.

**School F.** Items were chosen with consideration for difficulty level and content balance. Tests administered consisted of an entirely new item set. Five hundred and seventy six total students participated in this data collection at this school; specifically, 216 students were administered the K-1 test, 180 students were administered the 2-3 test, and 180 students were administered the 4-5 test.

**Results**

A summarization of the mean, standard deviation, minimum and maximum values for the three IRT parameters is presented in Table 3 for each reading domain. The number of items developed in each reading domain for the different ability levels are presented in Table 5. In total, 638 items from the CBAS-2 item bank were administered during the Fall and Winter of 2009-2010. Approximately 4,060 students participated in the study, allowing for approximately 300 responses per item. 16 items were selected and used as linking items, while 622 additional items were parameterized and retained for the subsequent version of CBAS-R, for use to index reading development from kindergarten through fifth grade.
References


Weiss personal communication, 2009
## Appendix

### Table 1

**School Demographics**

<table>
<thead>
<tr>
<th>Category</th>
<th>School A</th>
<th>School B</th>
<th>School C</th>
<th>School D</th>
<th>School E</th>
<th>School F</th>
<th>School G</th>
</tr>
</thead>
<tbody>
<tr>
<td>White</td>
<td>89%</td>
<td>68%</td>
<td>78%</td>
<td>72%</td>
<td>69%</td>
<td>76%</td>
<td>63%</td>
</tr>
<tr>
<td>Black</td>
<td>3%</td>
<td>16%</td>
<td>7%</td>
<td>7%</td>
<td>5%</td>
<td>4%</td>
<td>5%</td>
</tr>
<tr>
<td>Hispanic</td>
<td>3%</td>
<td>8%</td>
<td>8%</td>
<td>7%</td>
<td>6%</td>
<td>4%</td>
<td>26%</td>
</tr>
<tr>
<td>Asian</td>
<td>4%</td>
<td>8%</td>
<td>7%</td>
<td>13%</td>
<td>19%</td>
<td>15%</td>
<td>6%</td>
</tr>
<tr>
<td>American Indian</td>
<td>1%</td>
<td>&lt;1%</td>
<td>&lt;1%</td>
<td>1%</td>
<td>1%</td>
<td>1%</td>
<td>&lt;1%</td>
</tr>
<tr>
<td>Free/Reduced Lunch</td>
<td>21%</td>
<td>42%</td>
<td>37%</td>
<td>24%</td>
<td>14%</td>
<td>20%</td>
<td>41%</td>
</tr>
<tr>
<td>Limited English Proficient</td>
<td>1%</td>
<td>12%</td>
<td>9%</td>
<td>14%</td>
<td>14%</td>
<td>15%</td>
<td>28%</td>
</tr>
<tr>
<td>Special Education</td>
<td>10%</td>
<td>8%</td>
<td>8%</td>
<td>11%</td>
<td>10%</td>
<td>11%</td>
<td>13%</td>
</tr>
<tr>
<td>Total School Population</td>
<td>470</td>
<td>663</td>
<td>396</td>
<td>782</td>
<td>766</td>
<td>690</td>
<td>638</td>
</tr>
</tbody>
</table>

### Table 2

**Participants by School and Grade**

<table>
<thead>
<tr>
<th>Grade</th>
<th>School A</th>
<th>School B</th>
<th>School C</th>
<th>School D</th>
<th>School E</th>
<th>School F</th>
<th>School G</th>
</tr>
</thead>
<tbody>
<tr>
<td>K-1</td>
<td>100</td>
<td>203</td>
<td>128</td>
<td>243</td>
<td>278</td>
<td>216</td>
<td>215</td>
</tr>
<tr>
<td>2-3</td>
<td>130</td>
<td>238</td>
<td>123</td>
<td>255</td>
<td>258</td>
<td>180</td>
<td>174</td>
</tr>
<tr>
<td>4-5</td>
<td>150</td>
<td>212</td>
<td>114</td>
<td>243</td>
<td>238</td>
<td>180</td>
<td>182</td>
</tr>
<tr>
<td>School Total</td>
<td>380</td>
<td>653</td>
<td>365</td>
<td>741</td>
<td>774</td>
<td>576</td>
<td>571</td>
</tr>
</tbody>
</table>
Table 3

**Summarization of Parameter Estimates for Each Domain**

<table>
<thead>
<tr>
<th>Domain</th>
<th>N</th>
<th>Parameter a</th>
<th>Parameter b</th>
<th>Parameter c</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Mean</td>
<td>SD</td>
<td>Min</td>
</tr>
<tr>
<td>Concepts of Print</td>
<td>64</td>
<td>1.13</td>
<td>0.13</td>
<td>0.72</td>
</tr>
<tr>
<td>Phonological Awareness</td>
<td>70</td>
<td>1.04</td>
<td>0.09</td>
<td>0.80</td>
</tr>
<tr>
<td>Phonics</td>
<td>132</td>
<td>1.14</td>
<td>0.19</td>
<td>0.74</td>
</tr>
<tr>
<td>Vocabulary</td>
<td>213</td>
<td>1.10</td>
<td>0.10</td>
<td>0.71</td>
</tr>
<tr>
<td>Comprehension</td>
<td>143</td>
<td>1.16</td>
<td>0.17</td>
<td>0.88</td>
</tr>
<tr>
<td>Whole Bank</td>
<td>622</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 4

**Item Difficulty Information**

<table>
<thead>
<tr>
<th>Domains</th>
<th>(-3&lt;b&lt;=-2)</th>
<th>(-2&lt;b&lt;=-1)</th>
<th>(-1&lt;b &lt;=0)</th>
<th>(0&lt;b&lt;=1)</th>
<th>(1&lt;b&lt;=2)</th>
<th>(2&lt;b&lt;3)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Concepts of Print</td>
<td>15</td>
<td>16</td>
<td>18</td>
<td>9</td>
<td>6</td>
<td>0</td>
</tr>
<tr>
<td>Phonological Awareness</td>
<td>6</td>
<td>15</td>
<td>27</td>
<td>15</td>
<td>7</td>
<td>0</td>
</tr>
<tr>
<td>Phonics</td>
<td>15</td>
<td>32</td>
<td>31</td>
<td>29</td>
<td>22</td>
<td>3</td>
</tr>
<tr>
<td>Vocabulary</td>
<td>13</td>
<td>8</td>
<td>39</td>
<td>43</td>
<td>49</td>
<td>61</td>
</tr>
<tr>
<td>Comprehension</td>
<td>0</td>
<td>1</td>
<td>48</td>
<td>55</td>
<td>23</td>
<td>16</td>
</tr>
<tr>
<td>Whole Bank</td>
<td>49</td>
<td>72</td>
<td>164</td>
<td>151</td>
<td>107</td>
<td>77</td>
</tr>
</tbody>
</table>
Figure 1. Illustration of CBAS-R administration conditions. The multimedia features allow visual and auditory input. They also isolate the student from extraneous stimuli that might otherwise distract the test taker. The keyboard is hidden with a keyboard cover and students responses are recorded by pointing and clicking with a mouse.

Figure 2. Linking item Unique ID 11; Audio: “Click on the picture that shows the front of a book.”
Figure 3. Linking item Unique ID 12; Audio: “Click on the word where you start reading.”

Figure 4. Linking item Unique ID 24; Audio: “Click on the first letter of the word.”
The farther away from the equator a town is located, the more it experiences variation in hours of daylight as the season changes. For people living nearer the equator, then, summer days and winter days have approximately the same _____.

a. temperature  
b. location  
c. number  
d. length

Figure 5. Linking item Unique ID 35; Audio: “Choose the best answer.”

The audience applauded and cheered and goaded the performer. They could not contain themselves for they were witnessing a first in history. Just as they thought the performance was over, an eruption filled their ears with an excruciating noise.

What is the theme of this passage?

a. bewilderment  
b. fantasy  
c. enlightenment  
d. thrill

Figure 6. Linking item Unique ID 43; Audio: “Choose the best answer.”
Poolside, the water hit the tiles gently in the breeze. Sage and rosemary from the nearby herb garden swelled into the owner’s nose as a soft, warming sun settled on her skin.

What is the theme of this passage?

a. valor  
b. serenity  
c. engagement  
d. underwhelming

Figure 7. Linking item Unique ID 45; Audio: “Choose the best answer.”

The president of the company was particularly fed up with one employee who disrespectfully interrupted and criticized his superiors.

a. indolent  
b. awkward  
c. presumptuous  
d. new

Figure 8. Linking item Unique ID 57; Audio: “Choose the best answer.”
Barry was a lauded ________, well-known for his woodworking skills.

a. plumber  
b. artisan  
c. ceramist  
d. taxonomist

Figure 9. Linking item Unique ID 81; Audio: “Choose the best answer.”

To begin with our old friends. Franz was a tall lad, of sixteen now, a regular German, big, blond, and bookish, also very domestic, amiable, and musical.  
--“Little Men”, Louisa May Alcott

Choose the synonym for amiable.

a. angry  
b. interesting  
c. honest  
d. friendly

Figure 10. Linking item Unique ID 67; Audio: “Choose the best answer.”
On the way to the grocery store, our car began to sputter. Neither mom nor dad knew what the problem could be, so they pulled over to the side of the road.

What most likely happened next?

a. Mom and Dad got out of the car to inspect under the hood.
b. Mom called her mechanic friend.
c. We complained about the heat outside the car.
d. Mom and Dad suggested that they walk to the store.

Figure 11. Linking item Unique ID 94; Audio: “Choose the best answer.”

"Bright & Shine will make your teeth sparkle like never before. Your smile will look clean and glowing!"

This ad claims that the product will make your teeth _____.

a. bigger  
b. harder  
c. whiter  
d. sharper

Figure 12. Linking item Unique ID 104; Audio: “Choose the best answer.”
Figure 13. Linking item Unique ID 176; Audio: “Which word does not have the same middle sound?”

Figure 14. Linking item Unique ID 379; Audio: “Click on the word ‘big’.”
Figure 15. Linking item Unique ID 381; Audio: “Click on the word ‘girl’.”

Figure 16. Linking item Unique ID 395; Audio: “Click on the word ‘belong’.”
Figure 17. Linking item Unique ID 395; Audio: “Choose the best answer.”