Notes on Reading (Achievement) Measures

Prepared by Ann Bosma Smit
Updated February 24, 2005

A large number of both published and constructed measures were used to assess achievement both in reading and in pre-reading skills. Virtually all of these tasks involved decoding rather than comprehension of a passage. However, a measure listed under Language, the “Aluminum Passage Rewrite” task, administered in Session 12, requires the participant to read and understand a text, and then rewrite it. Inferences about comprehension can be made from this task.

Gates Reading Survey, 1960 Revision (Gates, 1960)
Administered at Session 11 (Spring 1965, when most participants were in grade 4)
Located in data boxes #11 and #12

The Gates Reading Survey includes three tests, as follows:

1. The Speed and Accuracy Test, which contains 36 paragraphs, each with a comprehension question or fill-in-the-blank. The response is chosen from among 4 foils. This is a timed test. The “speed of reading” score is the number of exercises correct, for a maximum of 36.

2. The Vocabulary Test, which has 65 items arranged in order of increasing difficulty. Each item presents a key word accompanied by five other words from which the child chooses the one whose meaning is nearest that of the key word (synonym). The score is determined by totaling the number of correct responses and subtracting one-fourth the number of wrong responses. The maximum possible score is 65.

3. The Level of Comprehension Test, which has 21 reading passages arranged in order of increasing difficulty, each with 2 or more blanks to be filled with words chosen from a short list (cloze procedure). The score on this section is determined by totaling the number of correct responses minus one half the number of wrong responses, for a maximum score of 43.

The survey itself can also be assigned a total score, which is the sum of the Speed, Vocabulary, and Comprehension scores, for a maximum of 144.

The three subtest scores and the total score are entered in columns 40-41, 42-43, 44-45, and 46-48, respectively, on card 1, Session 11 (Spring 1965, grade 4).
**Grade 1 Vocabulary Recognition (also called Scott-Foresman 1st Grade Word List)**  
Constructed by G. Jain and M. Templin  
Administered at Session 5 (Spring 1962, grade 1)  
Located in Data box #13  

This measure consisted of 106 words selected from the vocabulary in the Scott Foresman *First Reader, the New Basic Readers Curriculum Foundation Series* (Robinson, Monroe, & Artley, 1956). First proper names were eliminated, and then a random numbers technique was used to select the words. The list of words, printed in lower case, was given to the child, who was asked to read the words aloud. For each word, the examiner noted the child’s response as (1) word correctly read, (2) no attempt made to read the word, or (3) word read incorrectly, with the exact response noted.

The score is the number of words read correctly, with a maximum possible score of 106.

The score was entered in columns 54-56 on card 1, Session 05.
Grade 2 Vocabulary Recognition (also called Scott-Foresman 2nd Grade Word List)
Constructed by G. Jain and M. Templin
Administered at Session 7 (Spring 1963, grade 2)
Located in data box # 5

This task was essentially the same as the preceding, except that only 100 words were selected, and they came from Book Two, Part Two, of the same Scott Foresman reading series (Robinson, Monroe, & Artley, 1956) as above. Administration and scoring were similar to the Grade 1 task. The maximum possible score was 100.

The score was entered in columns 38-40 on card 1, Session 7.
Letter Sound-Production Test (also called Grapheme-Phoneme Production or Symbol-Sound Production)
Constructed by R. Johnson and M. Templin
Administered at Session 8 (Fall 1963, grade 3)
Located in Data box #21

Note: As described below, this task was considered by Templin to be a “Reading” task. A similar task was administered in Session 9, but that task involved listening to a sound and writing or pointing to a letter match, and it was considered to be a “Spelling” task. In addition, there appears to be a study of a subset of the children asking them to chose one of four choices in response to an auditory and a visual stimulus. The raw data are in Box 24.

In this task the child was asked to produce an acceptable phoneme in response to isolated capital letters of the English alphabet, along with a few digraphs (TH, SH, CH, WH, NG). The examiner instructed the child to tell him/her what sounds these letters make. The examiner transcribed the child’s responses using the International Phonetic Alphabet. A response was considered correct if it was a phoneme that could reasonably be associated in American English with the letter or letters on the list. The children’s responses were assigned codes, and the number of items assigned each code was tallied, for a possible total of 31. Consonants and vowels were not considered separately.

The tallies were entered in columns 42-53 (two columns per code) on card 1, Session 8.

Old (1963) Codes for Letter-Sound Production Task

<table>
<thead>
<tr>
<th>Code</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>Correct sound in isolation</td>
</tr>
<tr>
<td>1</td>
<td>Correct sound as the initial sound in a syllable or nonsense one-syllable word.</td>
</tr>
<tr>
<td>2</td>
<td>Correct sound as the final sound in a syllable.</td>
</tr>
<tr>
<td>3</td>
<td>Correct sound within a word. Frequently two syllable words, names, etc.</td>
</tr>
<tr>
<td></td>
<td>(appear to be associated with lack of understanding of instructions.</td>
</tr>
<tr>
<td></td>
<td>Often occurs with NR and letter names.)</td>
</tr>
<tr>
<td>4</td>
<td>Letter name (never use for vowels)</td>
</tr>
<tr>
<td>5</td>
<td>Incorrect sound in isolation</td>
</tr>
<tr>
<td>6</td>
<td>Incorrect sound in syllable—position disregarded</td>
</tr>
<tr>
<td>7</td>
<td>Incorrect sound in a word</td>
</tr>
<tr>
<td>8</td>
<td>No Response</td>
</tr>
<tr>
<td>9</td>
<td>Correct sounds of th, c, r, etc. both given, e.g. both voiced and voiceless th</td>
</tr>
<tr>
<td>11</td>
<td>One correct sound and one incorrect sound given</td>
</tr>
<tr>
<td>12</td>
<td>Two incorrect sounds given</td>
</tr>
</tbody>
</table>
(Untimed) Word Pronunciation Task
Constructed by R. Johnson
Administered at Session 8 (Fall 1963, grade 3)
Located in data box #21 and #24

This task was designed primarily to identify the technique(s) of word recognition that a child uses when attempting to pronounce what are probably unfamiliar words with no context. There were five polysyllabic words (PAPER, DOWNFALLEN, INTEREST, CONDUCTOR, and SUPERSTITION), each typed in all caps on a 3x5 card. The children were encouraged to make all attempts at pronunciation orally, and the examiner recorded all attempts for each word. After the child had attempted all five words, the examiner recorded his/her impression of the most prevalent characteristics of the child’s overall performance, e.g. “depends mainly on general appearance of a word”; “recognizes at sight familiar parts—root words, parts of compound words, etc.”; “analyzes syllables, phonograms or other meaningful units phonetically”; “depends mainly on sounding individual letters”; and “depends mainly on naming individual letters. In other words, the examiner performed a kind of miscue analysis.

Templin’s 1968 monograph indicates that the two quantitative scores were (1) the number of words pronounced correctly on the first attempt, for a maximum score of 5, and (2) the number of words pronounced correctly regardless of the number of attempts, maximum score of 5. However, the data entry guide of 8/72 provides for a total score and a number of process codes used to indicate the nature of the child’s word-attack schemes. These codes are:

1 = Depends mainly on general appearance of word
2 = Recognizes at sight familiar parts—root words, parts of compound words, etc.
3 = Analyze syllables, phonograms, or other meaningful units phonetically
4 = Depends mainly on sounding individual letters
5 = Depends mainly on naming the letters (spelling method)
6 = Does not blend well
7 = Lacks versatility; if first sounding or blending is wrong, does not work out others
8 = Trouble with consonants
9 = Trouble with vowels

On Card 1, Session 8, the number of words pronounced correctly and fluently on first attempt is entered into column 27. The number of words that are eventually pronounced correctly, even if non-fluently, is entered into column 80. The total number of process items checked is entered in column 28. If only one process item is checked, it is entered into Column 29. If two process items are checked, Templin implies in the test description that each item number is entered into columns 30-31. However, the database does not seem to work that way. A partial coding when there are two items is this (based on spot-checks of the data):
<table>
<thead>
<tr>
<th>Process item numbers</th>
<th>Code</th>
</tr>
</thead>
<tbody>
<tr>
<td>none</td>
<td>00</td>
</tr>
<tr>
<td>1 and 2</td>
<td>10</td>
</tr>
<tr>
<td>1 and 6</td>
<td>40</td>
</tr>
<tr>
<td>2 and 3</td>
<td>60</td>
</tr>
<tr>
<td>8 and 9</td>
<td>08</td>
</tr>
</tbody>
</table>

When three or more codes are checked, there is an arbitrary coding scheme that is entered into columns 32-35. That scheme is as follows (and some may be missing):

<table>
<thead>
<tr>
<th>Process item numbers</th>
<th>Code</th>
</tr>
</thead>
<tbody>
<tr>
<td>none</td>
<td>0000</td>
</tr>
<tr>
<td>5, 6, 7, 8</td>
<td>0001</td>
</tr>
<tr>
<td>5, 7, 8, 9</td>
<td>0003</td>
</tr>
<tr>
<td>6, 7, 8, 9</td>
<td>0004</td>
</tr>
<tr>
<td>1, 6, 7, 8, 9</td>
<td>0005</td>
</tr>
<tr>
<td>5, 6, 7, 8, 9</td>
<td>0006</td>
</tr>
<tr>
<td>any six codes</td>
<td>0007</td>
</tr>
<tr>
<td>4, 8, 9</td>
<td>0010</td>
</tr>
<tr>
<td>5, 8, 9</td>
<td>0020</td>
</tr>
<tr>
<td>1, 6, 7, 8</td>
<td>0030</td>
</tr>
<tr>
<td>1, 7, 8, 9</td>
<td>0050</td>
</tr>
<tr>
<td>2, 3, 6, 7</td>
<td>0060</td>
</tr>
<tr>
<td>2, 4, 8, 9</td>
<td>0070</td>
</tr>
<tr>
<td>4, 6, 7, 8</td>
<td>0080</td>
</tr>
<tr>
<td>4, 6, 8, 9</td>
<td>0090</td>
</tr>
<tr>
<td>1, 6, 7</td>
<td>0100</td>
</tr>
<tr>
<td>1, 6, 8</td>
<td>0200</td>
</tr>
<tr>
<td>1, 7, 8</td>
<td>0300</td>
</tr>
<tr>
<td>1, 8, 9</td>
<td>0400</td>
</tr>
<tr>
<td>2, 3, 6</td>
<td>0500</td>
</tr>
<tr>
<td>2, 3, 7</td>
<td>0600</td>
</tr>
<tr>
<td>2, 6, 8</td>
<td>0700</td>
</tr>
<tr>
<td>2, 8, 9</td>
<td>0800</td>
</tr>
<tr>
<td>1, 2, 3</td>
<td>1000</td>
</tr>
<tr>
<td>1, 2, 7</td>
<td>5000</td>
</tr>
<tr>
<td>1, 3, 5</td>
<td>6000</td>
</tr>
<tr>
<td>1, 4, 7</td>
<td>8000</td>
</tr>
</tbody>
</table>
This is actually Test IV, Sounding Whole Word Test, in the McCullough Word Analysis Tests, Experimental Edition (McCullough, 1962). This test gets at “the ability to identify consonant and vowel sounds and to blend them into whole words” (Manual, p. 2). This particular test was designed for children ready for the fifth-grade reader, but in the Templin study, it was administered in the fall of the year in which most of the participants would have been in the fourth grade. The test includes 30 items, each of which consists of three strange-looking “words” and a dashed line. None of these “words” are correctly spelled English words, but some of them could be English words if the child were aware of the sounds associated with the letters.

The examiner gave these instructions to the child: “This is a test of your ability to sound out strange words. Sound out to yourself the words in each row. If one of them sounds like a word you have heard and know the meaning of, put a cross on it. If no word in the row sounds like one you know, put a cross on the blank at the end of the row.” The examiner worked through a practice item with the child to be sure she understood the procedure.

The score is the number of correct items, with a maximum score of 30.

The score is entered in columns 36-37 on card 1, Session 10. The score is also on one of the background cards.
McCullough Syllabification Test
Administered at Session 10 (Fall 1964, grade 4)
Located in data box #18

This is actually Test VI, Dividing Words into Syllables Test, in the McCullough Word Analysis Tests, Experimental Edition (McCullough, 1962). The test has 30 two-syllable words, and the child’s task is to draw a line between the two syllables. Eight different rules for “syllabication” (syllabification) are represented, e.g. “Divide before a single medial consonant,” as in la | bor.

The score is the total number of words that are syllabified correctly, for a maximum of 30.

The score is entered in columns 38-39 on card 1, Session 10.
This is actually Test 7: Word Elements from the Silent Reading Diagnostic Tests by Bond, Clymer, and Hoyt (1955). [Note: Templin references this test as the Developmental Reading Tests by these authors.] It appears to be more of a spelling test than a reading test. This test contains 30 items. The child hears the examiner pronounce a syllable such as cen (as in the word center) and then picks out the appropriate spelled form from an array that might include con, kom, son, nec, and cen. (There is a template to assist in scoring this test.) The score is the total number of correct choices, with a maximum score of 30.

The score is in columns 40-41 of card 1, Session 8.
This test is *Test 4. Locating Elements* from the *Silent Reading Diagnostic Tests* by Bond, Clymer, and Hoyt (1955). [See note above.] The child sees a picture of a fairly short word, such as *fan*, and is to underline that short word in a longer word printed nearby, for example, *fancy*.

Two scores are recorded for the child, the first being the number of items completed correctly within 5 minutes, and the other being the number of items completed correctly with no time limit. The maximum score for both is 36.

These scores are entered in columns 36-37 and columns 38-39, respectively, on card 1, Session 8.
Gates Advanced Primary Reading Test—Word Recognition and Paragraph Reading  
(*Gates, 1958*)
Administered at Session 7 (Spring 1963, grade 2)
Located in data box #3

These tests were administered by classroom teachers as part of the regular testing program of the Minneapolis Public Schools, or by examiners to small groups and individuals in the parochial and public schools. Unfortunately, there were many participants who were missed, and so these data are excluded from most analyses.

The Advanced Primary Word Recognition Test requires the child to look at a picture and then circle one of four printed foils at the right of the picture to label the picture. For example, for a line drawing of a snail, the foils were:

- flail
- snail
- snake
- snore

The maximum score for the *Advanced Word Recognition (AWR)* is 48, and scores are corrected for guessing (total correct minus one-third the number of errors).

The Advance Primary Paragraph Reading task presents the child with several pictures. The child then reads the text under the drawings and carries out the printed instructions therein. The text ranges from short single sentences (e.g. “Draw a line around the milk bottle”) to a longer passage (e.g. for a drawing of a woodpecker on a tree trunk: “One of the most beneficial inhabitants of the forest is the woodpecker. With its feet, consisting of two toes in front and two behind, it is able to grip the bark of a tree firmly, and with its long bill it can drill into the bark for insects, which are its food. Draw a line under its unusual feet. Put an X on the thing with which it drills into bark.”) There are 24 panels on the test, some of which contain two instructions. Consequently, the maximum score for the *Advanced Paragraph Reading (APR)* is 32.

The AWR score was entered in columns 34-35 on card 1, Session 7. The APR score is entered in columns 36-37 on the same card.
Homographs, also called Phonemes in Sentences, also called Stress Shift
(Constructed by S. Rosen and M. Templin)
Administered at Session 11 (Spring, 1965, grade 4)
Located in data box #28

This test consists of 18 sentences for the child to read out loud. Each sentence was
printed on a card, and the child read each one as it was presented to him. The test had a
number of goals, as follows:
1. to determine the child’s ability to correctly stress homographs, in which stress
   is the only cue to a difference in part-of-speech and meaning, e.g. content and
   content.
2. to assess the level of difficulty in the oral reading of specific sentences, and
3. to assess the adequacy of articulation of /s l r/ in sentences. (This part of the
task is described under Articulation Measures.)

To score for Goal 1, the examiner noted correct or incorrect productions. He/she also
rated incorrect productions with respect to an elaborate scoring system related to shift of
accent, repetition or hesitancy. The maximum score was 18. There is a very extensive
coding scheme for the homograph section, relating to hesitancy, stress shifts, etc.

To score for Goal 2, the examiner rated each sentence as a whole using the following
scale:
0 = no difficulty
1 = slight difficulty
2 = considerable difficulty

These numbers were used as weights to determine a weighted average (total?) with a
maximum score of 36 (with higher scores indicating greater difficulty in reading).

For Goals 1 and 2, several different scores are entered in columns 30-60 of card 2.
Session 11.

The scoring for Goal 3 is described under the Articulation Measures as Phonemes in
Sentences.
**Gates MacGinitie Reading Test--Form F** (Gates and MacGinitie, 1969)
Administered at Session 12 (Spring 1972, grade 11)
Located in data box #36 (n = 374)

This test has four subscores: (1) Speed and accuracy of words attempted and words correct, (2) Vocabulary, and (3) Comprehension (using cloze-type procedures)

The subscores are entered in columns 58-67 on card 1, Session 12.
Woodcock Reading Mastery Tests, Norming Edition: Word Identification and Word Attack (also called the Rapid Reading Task; Woodcock, 1972)
Administered at Session 12 (Spring 1972, grade 11)
Data are located in data box #32

This test had a number of scores/subscores. Scores entered for these tests included a Total Score, Word Identification, Word Attack, Word Comprehension I (just called “Word Comprehension” in the test forms), and Word Comprehension II (called “Passage Comprehension” in the test forms). Both types of word comprehension involve cloze procedures, where the child reads something and chooses the best word to go in the blank space. Word Comprehension I is actually similar to the Miller Analogies Test. Passage Comprehension uses short and long passages with a word missing.

This test was apparently being normed at the time that Templin used it. The author supplied the distributions for each section of 220-plus children. Templin developed scaled scores based on these distributions. However, these scales represent neither standard scores nor percentile ranks, and in fact some of the scaled scores can have negative values.

The raw scores are entered into columns 58-67 on card 1, Session 12. The corresponding Scaled Scores are entered into columns 7 - 26 on card 3, Session 12.
Arthur, the Young Rat
Administered at Session 12 (Spring 1972, grade 11)
Located in Data Box #30

[NB: This is the only place I have found reference to dialect in an articulation scoring system. Also, I cannot find the source of the Arthur passage. It seems to sample the phoneme inventory of English.]

Arthur, the Young Rat, is a passage that the child reads out loud. The examiner records any errors, including both articulation and reading errors. It appears that this sample was not recorded on tape. The quantities and types of errors are coded with single digit codes. In addition, Templin used a series of binary (Yes/No) codes to describe the articulation and reading errors made by the child. Also included in the binary section are items that would be considered to relate to general speech ability, such as voice deviances, and so forth.

The codes relating to this reading passage are entered in columns 31-49 on card 1, Session 12 (Spring 1972, grade 11), one column per variable. The binary data are on card 2, Session 12, in 49-64.
References


