Framework for the Assessment of Reading Component Skills

John P. Sabatini
Educational Testing Service

Literacy Assessment and Monitoring Program (LAMP)

June 2004

Contents

I. Introduction

> 'What a child learns is how his or her writing system works – both its basic principles and the details of its orthographic implementation. We know this learning has occurred when the child can identify printed words as words in his or her spoken language in a way consistent with the writing system. For an alphabetic reader, this means being able to read unfamiliar words, and even nonwords, as well as familiar words. For a Chinese reader, this means identifying familiar characters and being able to make informed guesses about the pronunciation or meaning of unfamiliar characters using their compositional principles. (This does not include reading nonwords, which in Chinese is impossible, strictly speaking.) To be sure, much more is learned than how one's writing system encodes one's language. But this is the central learning even to which additional literacy learning, for example, comprehension strategies, must be connected.' (Perfetti, 2003, p. 16)

The Assessment Framework described in this report builds upon this basic principle of learning to read described by Perfetti, now widely researched and accepted internationally. That is, the comprehension or 'meaning construction' processes of reading are built upon a foundation of component skills and knowledge of how one's writing system works. The evidence of this knowledge and skills can be captured in tasks that examine a reader's ability and efficiency in processing the elements of the written language – letters/characters, words (and nonwords), sentences, and larger, continuous text segments.

A second principle guiding the components design is that the main interest is in whether the adults surveyed can apply their existing language and comprehension skills to the processing of printed texts. The components tasks are not designed to separately assess the level of language skills in the target print literacy writing system. It is assumed that the adults surveyed will have basic oral vocabulary, syntactic/ grammatical, and listening comprehension skills in the target language. We provide a component measure of basic oral vocabulary as an indicator that individuals surveyed have a threshold level of language proficiency. However, independent measurement of language proficiency is not a basic feature of the component framework.

A third principle of this model of reading is that the level of proficiency, efficiency, and integration of component skills is indicative of level and learning potential in reading development. As skills and knowledge accumulate, the ease of processing of familiar text-based information increases. Component efficiency is typically indexed by assessing speed or rate of processing, as well as accuracy. As learners, we spend extra time, effort, and energy to solve problems that are novel. On familiar tasks, we can often respond, accurately, quickly, with seemingly little conscious effort. When the tasks are easy, we can spend more effort solving and learning from more complex problems and tasks.

Finally, two guiding assumptions of this assessment framework are made. The first is that the adults to be sampled using the Component Measure LAMP tasks are in the low end of the continuum of reading ability (as evidenced by low performance on the screening instrument). The model of reading acquisition, development, and choice of item types and difficulties described below holds most strongly in this range of non- and developing readers. Different assumptions about component inter-relationships may hold for a population of more skilled readers.

The second assumption is that each country will *develop* a set of component measures unique to its language and culture based on the guidelines specified in the reading components framework and guidelines in Appendix A. As the relationship of the language to the writing system may be very different in different languages, the components and the nature of items and tasks to assess the component will need to be adapted based on consideration of those differences.

I.A. MEASURING COMPONENT SKILLS

The primary goal of the component skills battery is to help us better understand the "reading" profiles of adults at the low end of the literacy spectrum. In designing these measures in English, we can adopt the assumptions of the simple view of reading to organize our assessments to maximize useful profile information. As described by Hoover & Tunmer (1993): "the simple view makes two claims: first, that reading consists of word recognition and linguistic comprehension; and second, that while each of these components is necessary for reading, neither being sufficient in itself." (p.3) Word recognition is a stronger predictor of reading level in the early years of reading development. As word recognition becomes more fluent and automatized, listening comprehension becomes a stronger predictor of reading ability, though word recognition continues to contribute significant variance even in skilled readers (Gough & Walsh, 1991; Cunningham, Stanovich, & Wilson, 1990; McCormick, 1994). Strucker, Yamamoto, & Kirsch (2004) use a similar component framework when they describe *print components* (e.g., decoding accuracy and fluency) and *meaning components* (e.g., oral vocabulary).

We are concerned with five components that are indicative of reading acquisition levels and integrated reading ability:

a) alphanumeric perceptual knowledge and familiarity,
b) word recognition,
c) word knowledge (vocabulary),
d) sentence processing, and
e) passage fluency.

In skilled reading, these components are integrated to support literacy performance. During acquisition, they may be measured separately, with different profiles having implications for learning, instruction, and policy. The former two (a,b) are primarily print components. The latter three (c,d & e) are linguistic or language meaning components.

I.B. QUESTIONS ADDRESSED BY COMPONENT SKILLS

The component skill measures that make up reader profiles are intended to provide information to the following questions[1] (stated in terms of alphabetic languages):

What percentage of the population can:

1. recognize (or name) the letters of the alphabet of the writing system accurately?

1a. recognize (or name) the letters of the alphabet accurately, rapidly, and with ease?

2. recognize (or name) single-digit numbers of the writing system accurately?

2'. recognize (or name) single-digit numbers accurately, rapidly, and with ease?

3. recognize (or name) common words that appear frequently in print in the writing system accurately. These common words are expected to be in the listening/speaking lexicon/vocabulary of an individual who is a speaker of the target language.

3'. recognize (or name) common words that appear frequently in print and are likely in the listening/speaking vocabulary of the individual accurately, rapidly, and with ease.

4. recognize (or produce) plausible pronunciations of novel or pseudowords [2] by applying knowledge of the sight-to-sound correspondences of the writing system.

4'. recognize (or produce) plausible pronunciations of novel or pseudowords accurately, rapidly, and with ease.

5. know the meanings of common, everyday words in the language targeted for reading literacy.

6. process simple written sentences and apply language skills to comprehend.

6'. process simple written sentences and apply language skills to comprehend accurately, rapidly, and with ease.

7. process simple written passages and apply language skills to comprehend.

---

[1] These questions are stated in terms of alphabetic writing systems with specific reference to English. Throughout we provide rationales and design details that can be used to guide designs in character-based writing systems.

[2] Novel words are simply words that have meaning in the language, but may be uncommon and therefore not in the mental speaking/listening lexicon/vocabulary of the individual. Pseudowords are sequences of letters that could be pronounced based on the writing system, but are not words that would have a dictionary meaning in the target language.

7'. process simple written passages and apply language skills to comprehend accurately, rapidly, and with ease.

I. C. THE IMPORTANCE OF THE EFFICIENCY (') QUESTIONS

Most current models of reading development recognize the necessity of acquiring proficiency in lower level reading skills to enable development of increasing capabilities of higher level comprehension skills (Abadzi, 2003; Perfetti, 1986, 1992, 2003). Consequently, it is important to measure not only the accuracy level in performing component skills, but also gain some sense of the degree to which they are becoming efficient, effortless, automatized, internalized, routine. [3] Measuring speed or rate of response is a straightforward behavioral proxy for these cognitive constructs of automaticity and efficiency. Individuals skilled in a cognitive domain can perform routine processes more rapidly (maintaining high accuracy), than individuals who are not skilled. Simply put, it is easier (and quicker) to do what one knows how to do, and easier yet (and quicker yet) if one knows how to do something well. [4]

In reading, the range of skills and knowledge that can be performed with ease is an important indicator both of the reading level achieved and the prospective for future reading growth under different learning and experiential conditions. In the design of the components framework, specific timed exercises are included to measure these construct aspects.

I. D. LANGUAGE SKILLS IN READING LITERACY COMPONENT MEASURES

Depending on the country context, the target language of reading literacy may or may not be the first or dominant spoken language of the respondents. It is assumed for the purposes of this framework that there are not large populations of individuals whose reading literacy far exceeds their oral language skills in the target language. This would be true of highly educated individuals who might learn to read in a language other than their dominant spoken language, such as students preparing for foreign study. Typically, this group would also learn commensurate speaking/listening as well as reading/writing language skills, but there could be cases in which the reading/writing skill was significantly stronger. This population should score more highly on the screening battery and are not the primary target of the components framework.

The more prevalent case may be individuals whose reading literacy is limited, in part, because their oral language skills are limited. They may have limited oral vocabulary or limited knowledge and command of the syntax and grammar of the targeted language of

[3] Though there are technical debates in the literature over the precise operational definitions of these terms, there is near universal agreement that the skilled reader demonstrates behaviors that are captured by such terms. For the purposes of this report, we will use the term 'efficiency' as shorthand for this set of terms.

[4] This logic of this claim has it's limits in generalizeability across all cognitive and motor-perceptual domains of human activity, but its applicability to component reading skills has been repeatedly demonstrated.

reading literacy. One may be interested in distinguishing individuals with weak oral language skills from others.

As noted in Section 1, the primary focus of the components framework is on the application of existing language skills to the task of processing printed text. For oral language skills to be directly measured, respondents would be required to respond to orally presented stimuli (listening tasks) or be asked to respond orally (e.g., picture vocabulary task – see a picture and provide a word to name it). We provide a basic oral vocabulary task as an indicator of whether a basic threshold of language proficiency has been met by individuals in the survey. The claims or inferences about the component reading skills of individuals who fail to meet this threshold should be interpreted with respect to their limited oral proficiency.

Several other issues of language influencing the validity of measures should be noted. First, the clarity and speed of oral instructions and items should be carefully standardized. While audio recording of instructions and items is a useful technique for standardizing administration, listeners with weak language skills may be more sensitive to viewing the lips of the speaker, or having instructions/items repeated. Second, dialects, accents, and other language variations may make spoken/oral responses more difficult to score. Third, if the individual must struggle to produce spoken responses, this may influence speed/efficiency measurement. The risk to valid measurement of these factors needs to be weighed based on the country/region context of testing.

II. Framework of Reading Components

II. A. ALPHANUMERIC

Recognizing the alphabet is a core prerequisite of reading ability and growth. It remains a significant predictor of early reading acquisition in the U.S. (Adams, 1990). Part of this is explained by the obvious need of visual recognition of the letters to understand oral instruction. If an instructor asks the learner to find the word that begins with the letter ‘bee’, then the learner must identify the visual symbol from the auditory label. This is the most basic step of sight-to-sound correspondence – matching the letter name to the printed symbol and vice versa. Even observing that not all the letter names correspond to letter sounds (e.g., the letter name of ‘w’ is pronounced ‘double-you’) in English and that different languages have different names for letters (e.g., in German ‘b’ is pronounced ‘bay’) does not change the fundamental value of knowing this sometimes arbitrary set of associations.

However, just accurately being able to puzzle out the names of letters is not as indicative as having automatized this important symbol system. This latter skill serves as a foundation for a) benefiting from oral instructional settings, and b) focusing attention on higher level skills. In the context of a broad survey it is indicative of individuals’ experience with the writing system either through schooling or attempts at

reading print. From data collected in the U.S. and elsewhere, the rate of rapid naming of alphanumeric symbols remains moderately correlated with overall reading skill across developmental and adult levels (e.g., van den Bos, Zijlstra, & Spelberg, 2002; Sabatini, 2002)

Rapid naming of alphanumeric lists can be used as basic reading measures, as well as covariates for better understanding profiles or eliminating extraneous variance from inferences we might wish to make about subgroups. The main types of information provided by these tasks are as follows.

- Index of familiarity with basic perceptual codes in English. (Numbers and letters should be overlearned symbol systems and frequent exposure to them should result in efficient perceptual identification.)
- Index of baseline pronunciation rates (absent other oral speech samples). (Individuals speak at different rates.)

Rapid naming of letters and numbers have typically been administered separately. Letters are a slightly stronger predictor of reading than numbers, but they are generally more strongly correlated with each other than with overall reading. This pattern from the research literature holds because frequent exposure to printed texts in rich text settings (e.g., in schools, workplace) typically involves exposure to both letters and numbers in high frequencies. However, it is conceivable in some settings, such as communities with minimal printed materials available, that exposure and knowledge of printed digits is higher than for letters or vice versa. This highlights the value of be measuring both independently, rather than assuming one can be a proxy for the other. Therefore, we recommend both letter and digit recognition tasks.

II. B. VISUAL WORD RECOGNITION

All models of reading development recognize the centrality of rapid, automatic visual word recognition to reading ability (Abadzi, 2003; Adams, 1990; Perfetti, 1985). The visually presented printed real word (a spelling or orthographic representation in alphabetic languages) is transformed by the perceptual-cognitive system for processing into semantic (meaning) and phonological (sound-based) code systems. It is widely documented that the sound-based code is used in auditory working memory during the process of meaning construction or comprehension (Gathercole & Baddeley, 1993).

The semantic and phonological systems described for reading are the same cognitive systems used in language comprehension more generally. This forms the basis for the claim that visual word recognition ‘feeds’ the more general language processing system that also is used when listening to language or internal speech (Perfetti, 2003). Put another way, the goal of word recognition is to permit the individual to use the full extent of their language skills to comprehend as early in the cognitive processing of print as possible.

Without going into great detail on the mechanisms of word recognition (which are still under study in the psychological sciences), there are two basic behavioral skills that are indicative of proficiency in word recognition. First, is the accumulation of sight word knowledge of real words in the language. In English, one can identify a relatively smaller set of words that appear frequently in everyday texts, as compared to all words in the language that one might find in a dictionary. Most of these frequent printed words are words most skilled speakers of a language have in their speaking/listening lexicon/vocabulary. [5] Accurate and rapid recognition of frequent words is a strong index of word recognition efficiency and proficiency.

The second, more fundamental skill is decoding (also referred to as *word attack* or *cyphering*). This skill enables the generation of plausible pronunciations of printed words and conversely, plausible phonetic spellings of heard words. Decoding has been described as the fundamental word learning mechanism in alphabetic languages (Share, 1997), and therefore an essential component to measure directly. In alphabetic systems, decoding requires knowledge and skills in how lexical and sublexical sight-to-sound correspondences represent words in the language. Acquiring mastery of this skill is somewhat easier in languages in which the sight-to-sound correspondences are highly regular and predictable (e.g., German, Serbo-Croatian, Spanish, Turkish). With only a modest input of instruction, learners in these languages can often generate pronunciations for novel printed words and produce the correct pronunciation (i.e., the pronunciation that matches the typical spoken form in the language).

In languages with less regular correspondences, there are many alternate pronunciations for any given spelling (and vice versa), so more learning and instructional effort may be required to achieve proficiency. For example, the 'ou' vowel sound is pronounced differently in the English words 'could', 'though', 'thought', 'found'. In contrast, 'word', 'bird', 'heard', 'curd', 'nerd' all rhyme when pronounced, but the vowel sound is represented visually by different letters.

As noted, sight recognition of frequent printed words is a direct index of the accumulated knowledge of word reading. Several sources are available for getting an approximate list of these frequent words (e.g., Kucera and Francis, 1967). The ability of an individual to read a selected sample of such frequent, well known words without the benefit of passage context, is a useful index of how many words an individual can recognize in print. However, one cannot tell based solely on accuracy whether the words were processed more like sight words or decodable words. The distinction is one of degree as much as kind. A sight word is a printed word that has been seen often enough by the individual that it is recognized 'by sight', in contrast to a novel or pseudoword in which one must apply one's decoding knowledge of sight-to-sound correspondences to generate a pronunciation.

[5] This line of reasoning begins to get more complicated as one tries to categorize grammatical and morphological features in determining what counts as a word. For example, a dictionary will not list every verb tense as a separate meaning, though visually and auditorily they are different. In general, morphological, grammatical and syntactic variations across languages interact with word recognition in different degrees as well.

In skilled reading, both skills are necessary and applied rapidly, automatically, and strategically as needed. If one only measured decoding skill, one might have an estimate of the growth potential for learning real words, but under or overestimate accumulated knowledge of sight word knowledge directly necessary to reading and understanding printed texts. If one only measured sight word knowledge, then one might under or over estimate the growth potential. For example, low literate adults in the U.S. have been found to have sight word knowledge of frequent English words gleaned from years of formal schooling and exposure to printed text, that overestimates the decoding knowledge and skills they can apply to learning novel words. Though they have some functional literacy ability, their reading growth seems to be stunted by their slow progress in learning new sight words (Davidson & Strucker, 2002; Greenberg, Ehri, & Perin, 1997, 2003; Sabatini, 2003).

Therefore, in the components measures for word recognition in English and other alphabetic languages that are less regular, sight word and decoding tasks should be assessed separately.

II. C. WORD MEANING (VOCABULARY)

Very simply, a barrier to reading text would be not knowing the meaning of those printed words in one's listening or oral vocabulary. One can infer meanings of unknown words from the context (while reading or listening), but this typically produces provisional, uncertain, and incomplete words meanings – the understanding of which must be separately verified (e.g., checking definition in a dictionary).

In the component skills framework, the first goal is to determine whether individuals can recognize or decode words in the everyday listening lexicon of average adult speakers of the language. That is, the emphasis is on the everyday words of the language, rather than specialized technical or academic words that may be known by some but not most of the population. This would be the language used in the neighborhood or market. It would be the language of popular newspaper, radio, and television content.

The goal of the vocabulary component measure of this survey, then, is not so much to measure the full extent of individual's vocabulary knowledge, so much as determining whether their component reading skill levels reflect the proficiency of reading texts that they could otherwise understand in a listening context. As reading skill develops, one would expect a greater facility in learning new words from print whether or not one hears them used in oral language contexts. Decoding skill is critical to this word learning function of reading (for generating plausible pronunciations and storing a memory trace), as well as strong reading comprehension skills (for inferring meaning from context). However, these word learning skills are more evident after one has acquired sight recognition common words in text that are already in one's oral vocabulary.

Assessing vocabulary knowledge can be slippery. One would not want to confound reading vocabulary (i.e., word recognition) with oral vocabulary. Furthermore, words have multiple meanings, individuals can know partial meanings of words, and know the

meaning of a word only in a specific oral context. Many of these difficulties can be circumvented by using words that are concrete and visualizeable, which can then be made into picture vocabulary items. Respondents are shown line drawings depicting common things and then asked to provide the verbal label (e.g., book, chair, cat). Care must be taken to select items that are expected to be well known by most adults in the population.

II. D. SENTENCE PROCESSING

A variety of psychological studies of reading show that the sentence is a natural breakpoint in the reading of continuous text (e.g., Kintsch, 1998). A skilled reader will generally pause at the end of each sentence. A variety of operations are typically performed including encoding the propositions of the sentence, making anaphoric inferences, relating meaning units to background knowledge and ongoing models of the passage model, and deciding which meaning elements to hold in working memory. Thus, each sentence requires some syntactic and semantic processing. Several measurement focal points are possible, depending on the assumptions about the population and claims one is interested in making.

- If one can assume that the population has a basic command of the grammar and syntax of the language, then the emphasis will be on whether they can apply their language skills in the context of printed text.
- If one cannot, then one may also put emphasis on assessing their basic command of the grammar and syntax of the language.

By controlling the difficulty of the vocabulary in a sentence (words the individual can recognize and knows the meaning of), one can vary the grammatical complexity to get an indicator of the individual's proficiency at syntactic, grammatical, and semantic processing of prose.

In the component measure of sentence processing, sentence length and complexity is varied. The individual is asked to make a judgment of truth or falsity based on the content of the sentence, either in relation to a common knowledge about the world (see (Example 1) or based on the internal logic of the sentence (Example 2). For example, one could write items such as:

Example 1: "The sky is green" True or **False**
Example 2: "If a house is taller than a tree, then the tree is shorter than the house" **True** or False

The primary emphasis is on whether respondents can apply their existing language skills in a reading literacy context, not on higher level vocabulary, knowledge, or reasoning skills. At higher levels of reading, complex print can be easier to process than the comparable listening task, because one can review the printed text over and over, but the listening task may overload working memory. However, at the basic

reading level, we are interested in a more fundamental construction of meaning as a building block for more continuous text processing.

The efficiency version of this task is designed by creating sets of like items of modest difficulty, then asking individuals to complete as many as possible in a fixed time. A simple judgment (such as True / False), puts most emphasis on basic literal comprehension. Language is often used figuratively and metaphorically. Furthermore, the logical truth or falsity of basic facts in the empirical world can be slippery, as one can imagine exceptions to most absolutes. In the item design of sentences for this component, an attempt must be made to minimize ambiguity of meaning as much as is possible.

II. E. PASSAGE READING

Skilled reading is rapid, efficient, and fluent (silent or aloud). In this survey we consider fluency as more of an observable property that emerges from skilled reading. Fluency is indicating that visual word identification processes are efficiently feeding language processing systems (e.g. working memory) to produce outputs. The outputs do not necessarily imply the construction of meaning or comprehension as we commonly imagine it. For example, meaningless decodable visual word representations in a grammatically familiar text pattern can provide sufficient information to drive phonological and syntactic representations, resulting in fluently read gibberish. That is, skilled readers can read familiar texts somewhat fluently without attending to the meaning. However, fluency, defined as a relatively error free reading of a simple passage at a normal speaking rate, is a solid indicator of the integration of some basic component skills. On the other hand, breakdowns in accuracy, rate, or both, suggest difficulties in other subcomponents.

In this component measure we would like the participants to focus attention on comprehension as they read. The design for this component task uses a forced-choice cloze paradigm, that is, a choice is given between a word that correctly completes a sentence and an option that is incorrect. The incorrect item is meant to be obviously wrong to a reader with some basic comprehension skills. Distractors may be grammatically or semantically wrong. By giving the participant only a fixed amount of time to do the task, a measure of reading efficiency is assessed.

A relatively more skilled reader should be able to choose the correct response with minimal interruption of the fluent flow of passage reading. Therefore, for the more skilled beginning reader, the measurement is more on passage fluency and efficiency. This is because, the reader should be able to choose the correct word quickly, without much effort, and continue on reading at a normal rate. For the very low skilled beginning reader, the measurement falls more heavily on the reading comprehension construct.

II. F. ITEM/TASK DELIVERY AND SCORING Guidelines

II—F1 Group versus individual administration

These reading components assessments are designed to be administered individually.

II—F2 Response types/Data capture

There are six main options for data capture in this kind of assessment. Each has strengths and weaknesses for administration, scoring, choice and difficulty of test construction, and measurement integrity:

- Audio data capture
- Live scoring of oral responses
- Respondent paper and pencil (choice)
- Respondent paper and pencil (constructed)
- Administrator codes respondent response (paper and pencil choice)
- Administrator respondent response (paper and pencil constructed)

Although live scoring of oral responses will be used for the flipbook portion of these assessments, audio data capture is also required. In the efficiency measures of Sentence and Passage reading, we use a paper and pencil, forced choice task design, in which participants are required to circle the correct response.

II—F3 Naming, oral reading, and oral response

The most efficient task type for measuring the alphanumeric and word constructs and the type used in this set of assessments is a list naming task. In this task type, a list of items is presented visually and the respondent is asked to name the items aloud or simply read the sentences or passages presented to them. They may also be asked and answer orally presented questions about stimuli. Examples of this technique in other published tests include:

- Alphanumeric: Rapid digit and letter naming in the Comprehensive Test of Phonological Processing (*CTOPP*: Wagner, Torgesen, & Rashotte, 1999).
- Word Accuracy: *Letter-Word Identification and Word Attack* subtests of the Woodcock-Johnson III Tests of Achievement (*WJ*: Woodcock, McGrew, & Mather, 2001); Wide Range Achievement Test – 3 (WRAT-3; Wilkinson, 1993)
- Word efficiency: *Sight Word Efficiency* and *Phonemic Decoding Efficiency* in the Tests of Word Reading Efficiency (*TOWRE*: Torgesen, Wagner, & Rashotte, 1999),
- Vocabulary: Boston Naming test*, Picture Vocabulary* subtest of *WJIII*
- Gray Oral Reading Tests (GORT:); Passage Comprehension of WJIII

There are several **advantages** to this task type.

- the stimulus materials are easy to construct,
- the instructions are simple

- the tasks are easy to administer
- tests are quickly completed (a couple minutes each on average)
- efficiency and speed of response relatively easy to capture
- maximizes amount of construct relevant variance for amount of testing time

The respondent is faced with a list of items (numbers, letters, pictures, sentences, or passages) and asked to name/read them aloud. The task type is also well suited for estimating processing efficiency, as the respondent is given a fixed time to respond to a set (e.g., Name as many words as you can in 45 seconds; read as much of this passage as you can in one minute). For each construct, sets have been created that cover a specific range of item difficulty, to maximize interpretability of the individual's ability level. As noted, more skilled readers respond more quickly to items they know well, resulting in more items/sets reached in the former case, or shorter overall time to complete items/sets in the latter. The main disadvantage to oral response tasks is that scorers must be well trained to judge correct from incorrect oral responses. An audio recording of the test session is an efficient means for ensuring quality control in response judgment.

II—F4 Choice paradigms

F4a: Pointing versus marking (circle, checkmark, letter, etc.)

Choice paradigms are defined here as any response type indicating a correct answer. This includes both the respondent physically pointing or circling/marking a visual stimuli or scoresheet. Given the likelihood that a percentage of the respondents may not readily recognize the entire alphabet or digits, a simple marking or pointing system is strongly indicated. In the set of assessments designed here, pointing is the primary method used. This method requires that the administrator mark the responses in a separate data sheet. It also lends itself to flipbook stimuli with one item/set per page.

F4a: Number of distractors

For psychometric purposes, more distractors or choices per item reduce the guessing parameter. The smaller the guessing parameter, the fewer items that may be necessary to achieve a reliable score. For example, if all items are simple choices (True vs. False; Correct vs. Incorrect), then by chance an individual would be expected to score around 50%. Multiple distractors are most valuable when creation of quality individual items requires significant effort by test developers and when response time per item adds significant time to overall testing session.

In this framework, when choice is required we use simple choices for most task types. The rationale is as follows:

- it is relatively easy to generate large numbers of simple choice items in the domains of interest (sentences and short passages);

- items are designed to be relatively easy to respond to and should not take much thinking time. In fact, we are designing items that should be relatively effortless and quick to respond to for more skilled readers.
- multiple distractors typically require more problem solving time to evaluate all possible choices
- for low literate respondents, this problem solving among choices may be a significant additional processing burden in itself
- measures of speed or time are most valid when the time spent per item is in automatic, speeded response to known items; problem solving and evaluating distractors will weaken the reliability of efficiency measurement

We have designed items where performance is based most heavily on previously acquired reading component knowledge and skill. The individual either has the skill level to accurately respond to the item or not. There is not much additional information in the context of the item to spend on thinking about a response. Therefore, extra time spent thinking about a response is itself evidence suggesting weak underlying knowledge and skills.

III. Conclusion

Based on the constructs and design guidelines described in this document, a set of reading components assessments has been developed. This set includes the directions for administration, record sheet, scoring guide, and prototype of the flipbook. An audio recorder and a stopwatch are also required to administer these assessments.

IV. References

Abadzi, H. (2003). *Improving adult literacy outcomes: Lessons from cognitive research for developing countries.* Washington, DC: The World Bank.

Adams, M.J. (1990). *Beginning to Read: Thinking and learning about print.* Cambridge, MA: MIT Press.

Cunningham, A.E., Stanovich, K.E., & Wilson, M.R. (1990). Cognitive variation in in adult students differing in reading ability: Associations with vocabulary, general knowledge, and spelling. Jurnal of Educational Psychology, *83*, 264-274.

Davidson, R.K. & Strucker, J. (2002). Patterns of word recognition errors among adult basic education native and nonnative speakers of English. *Scientific Studies of Reading*, 6(3), 299-316.

Gathercole, S.E. & Baddeley, A.D. (1993). *Working memory and language*. Hillsdale, NJ: Erlbaum.

Goodglass, H., & Kaplan, E. (1983). Boston Diagnostic Aphasia Examination (BDAE). Philadelphia: Lea and Febiger.

Gough & Walsh (1991). Chinese, Phoenecians, and the orthographic cipher of English. In S. Brady & D. Shankweiler (Eds.), *Phonological processes in literacy* (pp. 199-209). Hillsdale, NJ: Erlbaum.

Greenberg, D., Ehri, L. C., & Perin, D. (1997). Are word-reading processes the same or different in adult literacy students and third-fifth graders matched for reading level? *Journal of Educational Psychology*, 89(2), 262-75.

Greenberg, D., Ehri, L. C., & Perin, D. (2002). Do adult literacy students make the same word-reading and spelling errors as children matched for word-reading age? *Scientific Studies of Reading*, 6(3), 221-241.

Hoover, W.A. & Tunmer, W.E. (1993). The components of reading. In G.G. Thompson, W.E. Tunmer, and T. Nicholson (Eds.), *Reading acquisition processes* (pp. 1-19). Philadelphia: Multilingual Matters.

Kintsch, W. 1998 *Comprehension: A paradigm for cognition*. Cambridge, UK: Cambridge University Press.

Kucera, H. & Francis, W.N. (1967). *Computational analysis of present-day American English*. Providence, RI: Brown University Press.

McCormick, S. (1994). A nonreader becomes a reader: A case study of literacy acquisition by a severely disabled reader. *Reading Research Quarterly, 29*, 156-177.

Perfetti, C. A. (1985). *Reading Ability*. New York: Oxford University Press.

Perfetti, C.A. (1992). The representation problem in reading acquisition. In P.B. Gough, L.C. Ehri, & R. Treiman (Eds.), *Reading acquisition* (pp.145-174). Hillsdale, NJ: Erlbaum.

Perfetti, C. (2003). The universal grammar of reading. *Scientific Studies of Reading*, 7(1), 3-24.

Sabatini, J.P. (2002). Efficiency in word reading of adults: Ability group comparisons. *Scientific Studies of Reading*, 6(3), 267-298.

Sabatini, J. P. (2003). Word reading processes in adult learners. In E. Assink & D. Sandra (Eds.), *Reading Complex Words: Cross-Language studies.* London: Kluwer Academic.

Share, D.L. (1995). Phonological reading and self-teaching: *Sine qua non* of reading acquisition. *Cognition*, 55, 151-218.

Strucker, J., Yamamoto, K., & Kirsch, I. (2004). In press.

Torgesen, J. K., Wagner, R. K. & Rashotte, C. A. (1999). *Test of word reading efficiency (TOWRE)*. Austin, TX: Pro-Ed.

Wagner, R. K., Torgesen, J. K., & Rashotte, C. A. (1999). Comprehensive Test of Phonological Processing (CTOPP). Austin, TX: Pro-Ed.

Wiederholt, J. L. & Bryant, B. R. (2001). Gray oral reading tests (GORT). Austin, TX: Pro-Ed.

Wilkinson, G. (1993). Wide Range Achievement Test 3. Wilmington, DE: Author.

Woodcock, R. W., McGrew, K. S., & Mather, N. (2001). Woodcock-Johnson III Tests of Achievement. Itasca. IL: Riverside Publishing.

van den Bos, K. P,. Zijlstra, B. J., & Spelberg, H.C. (2002). Life-Span data on continuous naming speeds of numbers, letters, colors, and pictured objects, and word-reading speed. *Scientific Studies of Reading*, 6 (1), 25-49.

V. Appendix A

General Guidelines for the Development of the Component Measures

NOTE: During the pilot phase for this battery, all test sessions should be audio recorded for scoring, item analysis, and quality control purposes. In any test development process, items must be piloted to make sure they behave appropriately in terms of participant response, scoring, and statistical characteristics. The pilot tasks include additional items in each task set. Items that are found to be flawed during the pilot phase or unnecessary for measurement needs can be eliminated for the full survey phase. We will evaluate at that time whether it would be necessary during the full survey to audiorecord for scoring or quality control purposes.

For all tasks except the Rapid Sentence and Passage tasks, we used a tented, flip book format. With this format, the administrator is facing the participant. The administrator can view the instructions and guidelines on his/her side of the tent book page. The participant can see a single page of stimuli at a time. After completion of a page, the administrator flips to the next page in the battery. The participant gives oral responses to flip book stimuli based on each task's instructions. The administrator marks correct/incorrect oral responses by the participant in a separate scoring booklet. For the Sentence and Passage Efficiency tasks, the participant marks responses directly on a task sheet.

A. Guidelines for Constructing the Numbers and Letter Naming tasks

A1. Digit Naming.

Use the numeric symbols for numbers from one to nine (1 to 9) and arrange in a random order in a single row. Use at least 14 pt type with wide line spacing. The administrator scores the task on the record sheet by drawing a line through each item incorrectly named or skipped.

A2. Rapid Digit Naming

Use the numeric symbols for numbers from one to nine (1 to 9) and arrange in a random order in five rows of nine digits each for a total of 45 items. Use at least 14 pt type with wide line spacing. Side conditions: no two of the same numbers next to each other; no more than three of the same number in a single row. Also make another sheet with two rows of eight items each. This is the practice list.

Based on previous research, we have set an initial time limit of 20 seconds. Pilot data should be used to confirm this time limit. Highly skilled readers can typically name all items in about that time, however, less skilled readers will likely not finish. The administrator scores the task on the record sheets by drawing a line through each item incorrectly named or skipped.

6 4 5 2 7 2 6 5 3

2 1 7 1 3 5 1 5 4

A3. Letter Naming

Arrange all the letters of the alphabet in rows. For the 26 letter, English alphabet, we used four rows of seven letters. Use at least 14 pt type with wide line spacing. Make sure every letter appears at least once. If there are extra spaces in the final row, repeat several letters to complete the row. Make two stimulus sets: one all upper-case letters, one all lower-case letters. The administrator scores the task on the record sheet by drawing a line through each item incorrectly named or skipped.

A4. Rapid Letter Naming

Arrange all the letters of the alphabet in rows. For the 26 letter, English alphabet, we used five rows with nine letters per row for a total of 45 items. Use at least 14 pt type with wide line spacing. Side conditions: no two of the same letters next to each other; no more than three of the same letter in a single row. Also make another sheet with two rows of eight items each. This is the practice list. We have used all lower case letters for the speeded list.

Based on previous research, we have set an initial time limit of 20 seconds. Pilot data should be used to confirm this time limit. Highly skilled readers can typically name all items in about that time, however, less skilled readers will likely not finish. The administrator scores the task on the record sheet by drawing a line through each item incorrectly named or skipped.

B. Guidelines for Constructing the Word Recognition and Decoding tasks

For all lists, we used lower case letters and at least 14 pt type with wide line spacing. Each list should be on a separate sheet.

B1. Real Word Lists

All words were selected from a print corpus database. Words with Kucera and Francis (K-F; 1967) frequencies of 100 words/million or better were identified. About 1000 of the most frequent words in print in English were identified in this way. Random numbers were generated for each word. Three classes of words were of interest: 1) 2-5 letter, single-syllable words, 2) 2-syllable, 4 to 5-letter words, and 3) 3-6 syllable words. These words were then combined to form all the lists according to the guidelines below. Proper nouns were eliminated.

The design goal is to construct lists in which the structure of words becomes progressively more complex, at the same time maintaining a relatively high word frequency and familiarity for a general population. Number of letters and syllables were the primary indices of complexity varied, as this has been repeatedly shown to be a chief indicator of word naming accuracy and rates. For single-syllable words, the consonant (C)-vowel (V) structure was also varied. In English, consonant clusters (e.g., CCVC as in ‘clam’) and vowel pairs (CVVC as in ‘main’) can be sources of difficulty in word recognition. Within lists, an attempt was made to separate any words that appeared closely related with respect to phonology, orthography, or semantics and

therefore might cause confusion for respondents or scorers (it/at, then/than; yes/no; more/most; bad/boy). The random selection is meant to allow for a variety of English spelling patterns including words with silent letters (e.g., same, light) or irregular pronunciations. Therefore, no attempt was made to control systematically for linguistic structure, regularity, or consistency of pronunciation. As these are all frequent words in print, the participant is expected to 'recognize' the word, not to 'decode' the sound pattern. For example, pronouncing 'have' as rhyming with 'save' would be considered an error, even though it is a reasonable decoding of the letter string. A practice list of 5 words and five lists of 9 words each were constructed.

- The practice list consists of five of the most frequent words in print in English (be, the, and, have, there).
- List 1 consists of two 2-letter words (VC), and seven 3-letter, 1-syllable words. Three words have CVC patterns, the others vary in CCV, CVV, and VCC patterns.
- List 2 consists of all 1-syllable, 4-letter words. The patterns CVCV, CCVC, CVCC, and CVVC, VCCV are represented.
- List 3 consists of all 1-syllable, 5-letter words. The patterns CVVCC, CCVCV, CVCCC, VVCCC, CCCVV, CVCCV are represented.
- List 4 consists of all 2-syllable words. The two 4-letter words have VCCV patterns. The seven 5-letter words have CVCVC, VCCVC, CVCCV, CVVCV, CVCVV patterns.
- List 5 consists of words ranging from 6 to 11 letters, and 2-4 syllables. In this list, some familiar morphological patterns (affixes and roots) are represented.

B2. Rapid Real Word Naming List

The same basic guidelines were applied in developing the rapid naming list as were used to develop the real word lists. All words were selected from a print corpus database. Words with Kucera and Francis (K-F; 1967) frequencies of 100 words/million or better were identified. About 1000 of the most frequent words in print in English were identified in this way. Random numbers were generated for each word. Three classes of words were of interest: 1) 2-5 letter, single-syllable words, 2) 2-syllable words, and 3) 3-4 syllable words. Although words become progressively more complex, we mixed some words. These words were combined to form a list of 60 items arranged in three 20 word columns.

- Column 1 consists of 2- to 4-letter, 1-syllable words. The list begins with 6 2-letter words. The next 4 words are a random mix of 2- and 3 letter words; the remainder of column one is a mix of 3- and 4-letter words. Some re-ordering of words was done when phonological/orthographic/semantic confusion may have resulted (e.g. in/is/it; bad/boy).
- Column 2 consists of 4- to 6-letter, 1- to 2-syllables words. Four and 5-letter words are randomly mixed in the first six items, the remainder are 6-letters in length.
- Column 3 consists of 6- to 10-letter, 2- to 4-syllables words. In this list, some familiar morphological patterns (affixes, and roots) are represented. The first 11

words are 6 to 7-letters, 2-3 syllable words, the remainder are 7- to 10-letters, and 3- to 4-syllable words.

Based on previous research, we have set an initial time limit of 60 seconds. Pilot data should be used to confirm this time limit. Highly skilled readers can typically name all items in about that time, however, less skilled readers will likely not finish. The administrator scores the task on the record sheet by drawing a line through each item incorrectly named or skipped and underlining the last word attempted before the time limit expired.

B3. Decoding Lists

Like in the real word lists, the design goal for the decoding task is to construct lists in which the structure of words becomes progressively more complex. Unlike on the real word lists in which words were meant to appear with high frequency in print and to be familiar for a general population, the decoding items should be novel. Participants must use their accumulated knowledge of English sight-to-sound patterns to generate pronunciations for spelling patterns that are novel, that is, not likely seen before or seen very infrequently.

There are three general strategies for designing items that require this knowledge and skill be applied. The first two strategies involve generating letter strings that can be pronounced, but are not real printed words in the English language (i.e., would appear in a dictionary). Spelling patterns that can be pronounced but would not sound like any English words we refer to as pseudowords (e.g., crimble, thist). Spelling patterns that sound like English words when pronounced are referred to as pseudohomophones (e.g., mait is pronounced like 'mate'). Finally, a third strategy is to select low frequency words or proper names that the participant is less likely to have ever seen in print (e.g., facund). The participant may or may not know the meaning of the word. They may or may not have even ever heard it pronounced. But they are unlikely to have seen it in print so as to recognize it as a sight word. In some cases, infrequent words are pseudohomophones of more frequent words (e.g., 'cite' appears less frequently in print than 'sight' or 'site').

Each strategy has strengths and weaknesses. Given the generally low reading literacy of this population, we expect a wider range of low frequency words that would not likely have been seen in print. We have chosen to use a combination of all three strategies. One chief advantage of pseudohomophones and low frequency words in a task such as this is that it is easier for the administrator to listen for a real word pronunciation (and perhaps one variation) and score it, than it is to listen for words never heard before (because they do not exist). Any specific participant may have an added source of information to aid in decoding (in the case of the pseudohomophone, does this printed word sound like a real word in my listening lexicon), but the task still requires decoding, as the specific item letter string presented cannot be recognized as a visual sight word (as it has not or rarely been seen in print before).

Infrequent words were selected from a print corpus database. Words with Kucera and Francis (K-F; 1967) frequencies of 10 words/million or less were identified. In the print corpus, proper nouns or word parts may appear occasionally in print, as well

as rare words. As noted, some infrequent words may be a second spelling of a more frequently seen word (e.g., prey/pray).

As with the real words, number of letters and syllables are the primary indices of complexity varied. For single-syllable words, the consonant (C)-vowel (V) structure was also varied. An attempt was made to control somewhat for regularity or consistency of pronunciation, however, this is not always possible because English has so many varying sight-to-sound patterns and so many borrowings from other languages. Nonetheless, every attempt was made to design items that had a single, most likely pronunciation given the structure of English spelling. As these are all invented or infrequent words in print, the participant is expected to 'decode' the sound pattern, not 'recognize' the word. A practice list of 6 words and five lists of 8 words each were constructed

- The practice list of 6 words consists of pseudohomophones (kar/car, fer/fur, hed/head, nale/nail, fite/fight, blud/blood).
- List 1 consists of 3-letter, 1-syllable words using all 3 strategies. Six words have CVC patterns, two CVV.
- List 2 consists of 4-letter, 1-2 syllable infrequent words. The patterns CVCV, CCVC, CVCC, and CVVC are represented.
- List 3 consists of all 1-syllable, 4-letter infrequent words. The patterns CCVC, CVCC, and CVVC are represented.
- List 4 consists of all 1-syllable, 5-letter pseudohomophones. The patterns CCVCV, CVCCC, CVCCV are represented.
- List 5 consists of 2-syllable, 5-letter pseudohomophones and infrequent words. The patterns CVCVC, CCVCV, CCVVC, CVCCC, CVCCV, VCCVC, CVVCV are represented.
- List 6 consists of 2-5 syllable, 5 to 11- letter pseudohomophones and infrequent words. In this list, some familiar morphological patterns (affixes, and roots) are represented.

B4. Rapid Decoding

The same basic guidelines were applied in developing the decoding lists. Three classes of words were of interest: 1) 2-5 letter, single-syllable words, 2) 2-syllable words, and 3) 3-4 syllable words. Although words become progressively more complex, we mixed some words. These words were combined to form a list of 42 items arranged in three 14-word columns.

- Column 1 consists of 3- to 4-letter, 1-syllable infrequent words. The list begins with 5 2-letter words. The next 5 words are a random mix of 2- and 3 letter words; the remainder of column one is a mix of 3- and 4-letter words. Some words re-ordering of words was done, when phonological/orthographic/semantic confusion may have resulted (e.g. in/is/it; bad/boy)
- Column 2 consists of 4- to 5-letter, 1- to 2-syllables words. Four and 5-letter words are randomly mixed in the first six items, the remainder are 5-letters in length.
- Column 3 consists of 6- to 10-letter, 2- to 3-syllables words. In this list, some familiar morphological patterns (affixes, and roots) are represented. The first 11

words are 6 to 7-letters, 2-3 syllable words, the remainder are 7- to 10-letters and 3-syllable words.

Based on previous research, we have set an initial time limit of 60 seconds. Pilot data should be used to confirm this time limit. Highly skilled readers can typically name all items in about that time, however, less skilled readers will likely not finish. The administrator scores the task on the record sheet by drawing a line through each item incorrectly named or skipped.

C. Guidelines for Constructing the Word Meaning (Vocabulary) Measure:

The word meaning (vocabulary) component has been constructed as a picture vocabulary exercise—that is, each item is a line drawing of an object; the respondent is shown the pictures one at a time and asked to give the name of the object. As the framework above describes in greater detail, this component is meant to assess the extent to which common, everyday words are in a respondent's lexicon or oral vocabulary. Items were selected, therefore, based on a combination of the following factors:

- high frequency in the Kucera-Francis (1967) word frequency list
- concreteness and imageability
- expectation that the objects are likely to be used or known by a speaker of the target language (i.e. the emphasis is on everyday vocabulary versus academic vocabulary)

To construct this assessment, a list of high frequency nouns was created using the Kucera-Francis list. These nouns then served as a general guide for the selection of the initial item pool. Although formal word frequency lists (like the Kucera-Francis) are useful in building item pools, one must also consider the imageability of the items. For instance, "day" is a high frequency word (686) and one that most speakers of a language could be expected to know. It is extremely difficult to create a line drawing that exemplifies "day," however. "Horse," on the other hand, is both frequent (117) and imageable. The line drawing of "horse" that appears as Item #1 of the assessment is an exemplar of the category. Respondents are unlikely to mistake the image for another animal or object unless they do not know the word in the target language.

The 20 items that comprise the assessment are ordered from most frequent to least frequent (117 to 1), are concrete nouns with high image ability (exemplars in their categories), and are likely to be in the lexicon of most speakers of the target language.

D. Guidelines for Constructing the Sentence Processing Items:

This component is comprised of 6 lists of True or False sentence items presented in the flipbook as well as 45 True or False sentence items presented as a paper-and-pencil

task on a separate sheet. The guidelines for constructing items are the same for all lists.

The task is to judge whether each sentence is true or false with respect to the ordinary world. The items were created so that the "truth" or "falsity" of the statements would be immediately obvious. None of the items should be ambiguous or vague, nor should they require any specialized background knowledge to answer correctly. For example, for the items "Eggs lay birds" and "Trees grow on the sun," general knowledge is sufficient to recognize immediately that these items are false.

Item difficulty is manipulated by varying word frequency, sentence length, and sentence structure. The list begins with short sentences comprised of high frequency (data obtained by consulting a frequency corpus) nouns and verbs using simple sentence structure: NOUN – VERB – DIRECT OBJECT. See examples below.

Table 1. Examples of some of the easiest items on the list.
Item: Water drinks people.
Item: Birds have feathers.

Noun	***Noun Frequency***	**Verb**	***Verb Frequency***	**Direct Object**	***Direct Object Frequency***
Water	*442*	drinks	*82*	people.	*847*
Birds	*31*	have	*3941*	feathers.	*6*

To increase the level of difficulty as the list progresses, we constructed items by using high frequency words in sentences of greater structural complexity--that is, by including noun phrases, verb phrases, and prepositions. See examples below.

Table 2. Examples of medium difficulty items.
Item: A person can run down a road.
Item: Food can be served on a table.

Noun Phrase	***Noun Frequency***	**Verb Phrase**	**Preposition**	**Object (+ article)**	***Object Frequency***
A person	175	can run	down	a road.	197
Food	147	can be served	on	a table.	198

Elements of sentence structure—specifically, the verb phrase and the preposition—drive most of the difficulty in the two items in Table 2. Word frequency does not drive the difficulty because all the words are expected to be in the oral vocabularies of the test takers. As is the case throughout the assessment, the correct answers to these medium difficulty items are obvious and immediate; item difficulty is not related to background knowledge.

To construct the most difficult items for the assessment, we used longer, more structurally complex sentences. See Table 3.

Table 3. Example of a higher difficulty item
Item: A person who has moved 100 miles has gone farther than a person who has moved 1 inch.

Noun Phrase	Relative Clause	Verb Phrase	Adverbial Phrase	2nd Noun Phrase	Relative Clause
A person	who has moved 100 miles	has gone	farther than	a person	who has moved 1 inch.

In the example given above, the length of the sentence has increased by including phrases and clauses within a comparative statement. While the correct answer is obvious, to reach it, the test taker has to process several pieces of information embedded within clauses and phrases. The item given above and ones similar to it are as difficult as the items get on this assessment.

Note on the placement of items: no more than three items with the same answer (either "true" or "false") are placed consecutively. Patterns of answers are also avoided (consistent "true," "false," "true," "false" patterning, for example). Approximately the same number of "true" and "false" items are embedded within the lists.

E. Guidelines for Constructing the Passage Fluency Items

This component is designed to assess the extent to which individuals can process simple written passages and apply their language skills to comprehend. The passages, therefore, are written in plain, simple language. They have minimal demands with respect to structure (syntax and grammar), vocabulary, and background knowledge. They are texts that would be understood by any adult speaker of the target language if the text were read aloud.

Because this is not an assessment of knowledge acquired during formal schooling, comprehension relies minimally on academic subject matter. Instead, the subjects of the passages were chosen to be familiar to almost any adult speaker of the language. Although it is impossible to pick subjects and write passages that are guaranteed to be understandable to people in all language groups and cultures, an attempt was made to make the passages as general as possible. For example, in this assessment, the subjects of the passages include food, transportation, and both urban and rural situations.

The task requires that the respondent read the sentences and choose between one of two words (the target word and a distractor) at the end of each sentence. The respondent is instructed to say or circle the word at the end of the sentence that makes the sentence make the most sense. In this regard, the distractors are designed to be obviously wrong syntactically and/or semantically. The distractors are also *high frequency*, obviously wrong choices to minimize the possibility that the respondent has never encountered one of the choices before. Local context should be enough for the respondent to choose the correct word—that is, the respondent should not have to

remember information from previous sentences in the passage or bring outside knowledge to bear in order to choose the correct word. The sentences themselves are not structurally complex, though introductory phrases and modifiers are used to create transitions and smooth prose. Total passage lengths are between 100 and 150 words.

Further, the task has been designed as both an oral and a silent reading exercise. Respondents will begin by reading aloud two sentences and a short, related passage. The final three passages have been designed as a paper-and-pencil, silent reading exercise; respondents will be given a fixed time to complete each silent reading passage.