2010

Assessing second language students' silent reading ability: Problems, perils, and promising directions

Gary Ockey, *Kanda University of International Studies*
D. Ray Reutzel, *Utah State University*

Assessing English Learners' Silent Reading Ability: Problems, Perils, and Promising Directions

Gary J. Ockey

Kanda University of International Studies, Japan

D. Ray Reutzel

Utah State University

The number of English Learners (ELs) in U.S. schools has increased dramatically over the last several decades (Montero & Kuhn, 2009). ELs struggle to attain high levels of literacy in U.S. schools for a variety of reasons. Failure to attain high levels of literacy often results in high dropout rates, unemployment or underemployment, and poverty for ELs (Kindler, 2002; National Center for Education Statistics, 2004). As a result of these factors, an increased focus on understanding and assessing the literacy development of ELs is urgently needed.

The report of the National Literacy Panel on Language-Minority Children and Youth (August & Shanahan, 2006) represents one of the first comprehensive efforts in the United States to collect and synthesize the findings of research focused on how second-language learners become readers and writers of English. One of the unfortunate outcomes, depending on one's point of view, found within the pages of this report was the persistent and general admission that the empirical research base in relation to developing ELs' literacy in English is fragile in terms of both quality and quantity. A more positive view on this lack of an empirical research base would be that there is plenty of room for more research on how to effectively develop ELs' English print literacy.

Complicating Factors in Assessing ELs' Silent Reading

Research suggests that ELs' reading ability depends to some extent on their general language proficiency (Brisbois, 1995; Taillefer, 1996; Yamashita, 2002). ELs with limited lexical, grammatical, textual, sociolinguistic, or other language abilities, normally brought to a silent reading task by primary or native language speakers

(L1s), may use additional or somewhat different cognitive and metacognitive processes. For instance, beginning L1 readers are generally faced with the task of decoding words that are already a part of their active vocabulary, while ELs more frequently encounter words they do not know. Proctor, Carlo, August, and Snow (2005) indicate that ELs often have smaller oral English vocabularies, which have been shown repeatedly to negatively affect reading comprehension.

To further complicate the matter, EL (as well as L1) beginning readers' general English language abilities are vastly different, ranging from completely proficient to absolute beginner. On one end of the continuum are students who are proficient English speakers when they begin to read. These students, like L1 beginning readers, have typically acquired English through oral interaction and, consequently, bring a great deal of English language proficiency to the task of silent reading. At the other end of the continuum are ELs who have no general English ability. In fact, many of these students begin communicating orally and reading English simultaneously. It is unlikely that the reading acquisition processes of students toward this end of the continuum are the same as those of L1s (or ELs) who learn to read after having acquired a fair amount of English language ability through oral communication. In addition, some ELs are literate in their primary language while others are not, and research suggests that some L1 reading strategies may transfer to the second language, especially for students who are somewhat proficient in English (Brisbois, 1995; Taillefer, 1996; Yamashita, 2002). Recent longitudinal research results, however, seem to suggest that L1s and ELs are more alike than different when acquiring English print literacy in the beginning stages (Fitzgerald, Amendum, & Guthrie, 2008). However, in later grades where reading comprehension, academic language, and content texts dominate literacy instruction, ELs tend to struggle (August & Shanahan, 2006).

Sorting out the reasons why many ELs struggle to attain print literacy levels comparable to their native speaking peers (L1s) is a persistent problem in literacy assessment (Garcia & DeNicolo, 2009). Questions about accurately assessing ELs' literacy attainment levels often involve several interrelated issues. First, for many ELs the presumed transfer of print literacy in the primary language to print literacy in English is not always reliable, because print literacy levels are often also low or nonexistent in the ELs' primary language. Second, literacy achievement tests are seldom norm referenced to a predominantly EL population (Butler & Stevens, 2001). Low levels of English print literacy can also result when tests of English print literacy are given to ELs who have low oral English skills (Butler & Stevens, 2001; Helman, 2009). Finally, sorting out learning disabilities from first- and second-language learning issues when testing ELs' print literacy in English is an even thornier issue (August & Shanahan, 2006). As a consequence, issues associated with assessing silent reading among ELs will likely be compounded and confounded by these factors in addition to known difficulties associated with assessing silent reading in the L1 population.

ELs face other challenges to their development of English print literacy. Many ELs live below the poverty level and, as a consequence, have access to few books

of interest or books written in their native language available in their homes or in school and classroom libraries. In addition, many books available to ELs in classrooms are too difficult for the majority of students to read (see Chapter 14, this volume). Martínez-Roldán and López-Robertson (1999) found that Latina/o students enjoyed books in which they saw people like themselves and in which their native language, Spanish, was represented. On the other hand, Mohr (2003) found that 84% of first-grade students, including Latina/o and African American students preferred English information books to a vast array of culturally relevant picture books they could have selected to keep. Au (2009) found that Hawaiian and Southeast Asian students neither knew how to find books in libraries nor how to choose books from those accessible to them. Perhaps one of the most encouraging findings from recent research on ELs' silent reading habits suggests that when ELs access interesting, culturally relevant, and appropriately challenging books and then have the opportunity to read daily in school with accountability, attitudes and reading habits are positively affected as well as end-of-year achievement test results (see Chapter 14, this volume).

A search for information about how to effectively assess ELs' English literacy acquisition results in a pronounced lack of empirical research evidence. It also is clear from this search that more federal investment in research that investigates the reliability, validity, and fairness of standards-based and norm-referenced literacy achievement assessments is clearly needed (Garcia & DeNicolo, 2009). Given the dearth of information about how to effectively assess ELs' English literacy acquisition, it should come as no surprise that a review of research on how to specifically measure ELs' silent reading performance would turn up a similar paucity of research evidence. Consequently, we rely not only on research addressing the assessment of silent reading among ELs but also on L1s as we discuss the potential problems, perils, and promises related to the assessment of ELs' silent reading processes.

What We Know About ELs' Silent Reading

For many years, silent reading was assumed to be one of the most effective practices for promoting ELs' English print literacy (Freeman & Freeman, 2008). This assumption was based largely on the work of Elley and Mangubhai (1983). These researchers reported a book flood project in which Fijian students read high-interest storybooks written in a second language. Volume reading of books in the second language led to reading growth rates twice that of students who did not read such books. This finding was interpreted into the widespread practice of engaging ELs in high-volume, independent, silent reading of English print materials as a means to increase their reading acquisition as the previous research had promised (Elley, 2000; Freeman & Freeman, 2008; Herrell & Jordan, 2004; Pilgreen, 2000). This presumption of independent, silent reading as effective practice held sway among teachers of ELs until the release of the National Reading Panel's (NRP; National Institute of Child Health and Human Development [NICHD], 2000) report, detailing the results of a research review and analysis. The NRP found the

existing body of evidence for independent, silent reading practice among L1s to be lacking in quality and quantity as well as failing to converge on a consistent finding that practice of this type yielded better reading achievement or motivation despite vehement claims to the contrary (Allington, 2002; Krashen, 2002). As a result, independent, silent reading as found in programs like sustained silent reading (SSR) or Drop Everything and Read was abruptly halted in many classrooms across the United States. The sudden cessation of independent, silent reading practice also blunted current and future efforts to seriously understand how one could assess the effectiveness of independent, silent reading on L1 or EL reading acquisition and motivation.

Independent, silent reading, although now making a comeback in many classrooms with recent research reports showing some convergence on the conditions to support effective silent reading practice (Kelley & Clausen-Grace, 2006; Kamil, 2008; Reutzel, Jones, Fawson, & Smith, 2008), remains elusive in terms of measurement for at least three reasons. First, measurement of the effectiveness of independent, silent reading practice on ELs' (and L1s') literacy acquisition or motivation remains elusive because silent reading practice is a cognitive activity not fully amenable to behavioral observation alone. Even though students' external reading behaviors, such as lip movement or vocalizations, may appear to be appropriate or inappropriate (see Chapter 4, this volume), there is no assurance that the necessary cognitive processes in reading are being employed by the student as he or she sits semiquietly in the corner with book in hand. Gaining a window into the silent, private cognitive processes of readers has long vexed those who would assess students' silent reading performance. Past attempts at measuring ELs' (and L1s') silent reading have centered almost exclusively on one of four assessment approaches or their variants: (1) asking students to read aloud brief portions of that which they have read silently; (2) having students complete self-reports of perceived reading ability and motivation, such as interest surveys, motivation surveys, or inventories before or after silent reading; (3) asking students probed and free recall questions about what they have read; and (4) asking students to retell what they have read silently. Recent variants of these approaches include use of computer technology and test accommodations aimed at increasing their validity.

A Survey of Assessment Problems and Perils With ELs' Silent Reading

Using Brief Oral Reading Events to Assess SSR

In the first of these typical assessment paradigms, students are asked to read aloud a short excerpt from their silent reading. Typically, the aim of the assessment is to establish that oral reading accuracy rates are sufficiently high, 95% or above, so that students are reading at their independent level, because little or no help or scaffolding with word-reading accuracy is available when reading silently (Stahl & Heubach, 2006).

A typical silent reading assessment may require a student to read aloud for one minute or less. This approach for measuring ELs' silent reading is inherently unsatisfactory because the task is not authentic; one cannot assume that a short, oral reading of a lengthy text is likely to be representative of students' accuracy or reading rate when engaged in lengthy periods of silent text reading. Although admittedly similar in cognitive demands, reading aloud and silently are different tasks. Because of these differences, one must question the authenticity of measuring the efficacy of silent reading with oral reading (see Chapter 9, this volume). Authenticity requires that to the extent possible, the characteristics of the test task (oral reading) must match the characteristics of the target language-use situation (SSR). The match between the tasks is crucial because it is an important determiner of the extent to which the test results can validly predict what test takers can do in the target language-use situation (Bachman & Palmer, 1996).

When silent reading ability is assessed by having students read aloud, the actual ability to read silently is confounded with oral abilities such as pronunciation and other verbal skills unnecessary for reading comprehension (Alderson, 2000). For instance, an EL may comprehend the text and read it accurately, but because of inaccurate or inarticulate pronunciation, an examiner may assume a lack of reading accuracy and assign the student an invalid score. Moreover, because ELs may have greater difficulty in articulating English words clearly than do L1s, they may demonstrate slower oral reading rates and, consequently, receive lower scores than their L1 silent reading ability counterparts. This is especially true in the earliest stages of English-language acquisition before the EL acquires English phonemic awareness and subsequently develops the ability to orally articulate unfamiliar phonemes in English.

Furthermore, a failure to read aloud accurately does not necessarily imply a lack of reading comprehension during silent reading. Occasional miscues, or deviations from what is written on the page when reading aloud, occur among even very skilled readers (Wallace, 1992). Goodman and Gollasch (1980) contend that miscues may be important for a reader's pursuit of understanding a text. For instance, some miscues may be used to reduce redundancy in text without changing the intended meaning. Miscues, when they are not driven by cognitive and metacognitive processes, have been generally viewed as reading errors in reading aloud assessments when, in fact, these miscues or deviations may be the product of lucid text comprehension in fluent readers (Goodman & Goodman, 1994).

Assessing ELs' silent reading with one-minute oral reading probes may also be problematic, because ELs may not be able to maintain the same reading rate for 15 minutes that they can sustain for a single minute. Some scholars refer to this phenomenon as reading stamina. Much like running, one's rate for running the quarter mile is quicker than one's rate for completing a marathon. As with runners, reading stamina relates to readers' ability to read longer texts while gradually increasing their rate. Because we do not know much about how reading stamina functions when students read silently, it is likely that the continued prevalent use of one-minute oral reading assessments used to calibrate current oral reading rate

norms fail to accurately estimate students' silent reading rates when reading for sustained periods of time in longer connected texts.

Students' reading stamina can be influenced by a variety of potential factors to include the amount of time and the length of text students are asked to read. Whether students struggle with word decoding or comprehension, which both make silent reading more effortful, the amount of effort required increases with the amount of time and the length of text to be read resulting in decreased reading stamina. Effortful word calling or a failure to understand what is read may be well tolerated in brief, one-minute oral reading assessment situations, but reading silently for longer periods of time is likely to lead ELs to disengagement from reading. Once disengaged, these readers often pretend to read silently or out of boredom engage in off-task behaviors, such as disrupting other students and wandering around the room. Thus, students who are assessed with short, oral reading assessments are motivated to perform at levels they might not otherwise choose to perform to please a teacher or to avoid appearing incompetent. Students who are placed in brief, oral reading testing situations experience different motivations to read than when they read independently and silently in books. One such example is the practice of teacher–student conferencing, in which students read aloud for one minute a small portion of the text they have been reading silently to assess the effectiveness of independent, silent reading practice. When reading silently for extended periods of time, those motivating conditions typically associated with the brief, oral reading testing situation alongside a teacher are not present.

The inauthentic use of brief, oral reading assessments to estimate ELs' silent reading accuracy, rate, or comprehension likely leads to erroneous conclusions about students' true silent reading rate or reading stamina. In addition, the inauthentic use of brief, oral reading assessments to estimate silent reading imposes on readers a different set of motivational conditions leading to different levels of engagement, or reading stamina, than what is associated typically with independent, silent reading.

Using Self-Report to Assess Silent Reading Ability and Motivation

Another approach for measuring ELs' (and L1s') silent reading is to have these students complete self-assessments of their perceived reading ability or an interest or motivation survey instrument prior to or after reading. Self-assessments typically ask students to indicate what they can read and understand. For example, DIALANG (Alderson & Huhta, 2005), a Computer Based Testing (CBT) system designed to assess second-language reading ability (as well as other second-language skills), contains a reading self-assessment (DIALANG is discussed in more detail later in this chapter). DIALANG uses "can-do" statements to obtain students' perceptions of their EL reading abilities. For instance, an example can-do statement from Level A1, a low level, is "I can understand very short, simple texts, putting together familiar names, words and basic phrases, by, for example,

re-reading parts of the text." An example of a can-do statement from Level C1, a high-level ability, states, "I can understand in detail a wide range of long, complex texts of different types provided I can re-read difficult sections." Although self-assessment of EL reading abilities has been shown to correlate highly (0.7) with objective tests of the same abilities (Ross, 1998), it is not clear that either the objective tests or the self-assessments are valid indicators of ELs' second-language reading abilities. Personality, mood, and other factors may all contribute to students' perceptions of second-language reading ability. Hiebert, Wilson, and Trainin (see Chapter 9, this volume) also note that self-reports of silent reading rates are problematic, because students may "fudge" the results when noting how many words they actually read during a timed reading, and they could make inadvertent mistakes during silent reading, such as skipping words, lines, or sections, resulting in skewed silent reading rate reporting.

The relationship between student motivation to read and reading stamina or engagement has been established for more than a decade (Guthrie & Wigfield, 1997). As with many correlations, however, we do not know if sustained reading generates interest and motivation for reading, interest and motivation generate sustained reading, or a third variable is the cause of both. To determine causality, there is a need for empirical studies that untangle the cause and effect of this relationship.

Using Probed and Free Recall to Assess Silent Reading

Using probed and free recall to evaluate silent reading comprehension presents yet another set of problems and perils when assessing ELs. In probed recall assessments of reading comprehension, students are asked questions about what was read silently. In free recall assessments of reading comprehension, the students are asked to give an oral retelling of what was read silently. A large number of task types have been used as probed and free recall assessments of silent reading comprehension.

Although the focus here is on ELs, probed and free recall, with few differences, are also used to assess L1s' silent reading abilities. One of the first probed recall tasks used to assess silent reading ability was the multiple-choice (MC) test developed by Kelly (1915). Kelly's silent reading test, the first published multiple-choice test for any purpose, used short stems (typically one or two sentences) along with a few possible answer choices, one or more of which were correct, and one or more of which were incorrect, to assess a students' silent reading ability (Barnwell, 1996). This MC format (with minor variations) has continued to be used for this purpose (as well as many other purposes) for nearly a century. Kelly's purpose in designing the MC format was to make large-scale testing more practical; the format made it possible for nonassessment experts to reliably and quickly score large numbers of tests. MC tests continue to be used widely as formats for assessing silent reading ability, because they remain practical, now even more so than when

they were developed by Kelly, because of the ability of computers to process and score such tests quickly and economically.

Over the past few decades, however, there has been mounting criticism of MC tests. Critics argue that they do not produce valid measures of reading ability, because the task is not authentic. Most readers do not interrogate themselves with questions during or after silent reading nor does comprehension of a text involve looking at answer choices after reading and attempting to identify the one which most associates information in the passage to information in a question stem. Moreover, many test takers do not follow their typical silent reading processes when taking MC tests, as is evinced by the reams of readily available published study guides designed to help students pass MC tests. These materials recommend such test-taking strategies as reading questions first and then scanning the text for answers. Given the lack of correspondence between students' typical silent reading cognitive processes and those employed when trying to achieve a high score on a MC test, it is highly unlikely that a score obtained from this probed recall assessment task type is a valid indicator of students' silent reading ability.

Other probed recall tasks that require students to respond to questions for assessing reading comprehension suffer similar shortcomings as the MC format. Alderson (2000) discusses a number of tasks designed to assess what readers comprehend. In matching tasks, students are given two lists and asked to match items on one with related items on the other. For instance, short passages might be matched with possible titles for the passages. Ordering tasks require students to place words, sentences, or paragraphs in an appropriate order. Some tasks designed to assess reading require students to use the additional skill of writing and therefore have limited validity as tests of silent reading ability. Short-answer tests require students to write responses to questions related to a text, and summary tests require students to read a text and then write a summary of its main points. Such tasks have been shown to assess writing ability as well as reading comprehension (Imao, 2008).

Using Oral Story Retellings to Assess Silent Reading

Story oral retelling, or free recall, is another popular task used for assessing ELs' silent reading ability, especially among young language learners. After a recorded story is played or read by a test administer, students are expected to orally retell the story, usually in as much detail as possible. As with brief oral reading events, inaccurate or inarticulate oral pronunciation may decrease the validity of the assessment. Moreover, effective completion of free recall tasks relies heavily on memory for text. Students may be able to comprehend what they have read but may forget to include some details when retelling the story. On the other hand, some students may be able to recite a text in detail without comprehending its meaning. Students who anticipate the requirement of producing a free recall or retelling a text following silent reading may employ strategies that are not typically employed during SSR. It follows, then, that the priming effect of knowing that a free recall is

required after reading a text may not yield tests scores that are valid indicators of students' authentic silent reading processes or products.

Using Technology to Aid in Assessing Silent Reading

The rather recent advent of CBT has affected the way silent reading ability is assessed. Early researchers believed that CBT had the potential to greatly increase the validity of assessment tasks (Green, 1983). CBT technology has made it possible to better control how assessment tasks are delivered to students and the processes needed to complete them. For instance, an experimental task used in DIALANG (Alderson & Huhta, 2005) is mapping and flowcharting. After test takers read a text and click on "continue," they select words from a list to drag into a map or a flowchart. Test takers can return to the text but cannot see both the text and the map or flowchart simultaneously. To some extent, the computer controls the strategies and processes the test taker uses to complete this task, because test takers are unable to see both the passage and the map or flowchart at the same time. Unfortunately, CBT-delivered tasks such as this one may be worse than their paper-and-pencil equivalents. Prohibiting test takers from viewing both the text and the flowchart simultaneously may further alter the processes a student might employ when reading silently. The CBT version of this test task may require more dependence on memory and less on comprehension and recognition abilities thought to be more closely tied to typical silent reading processes (Ockey, 2009). Lack of access to or familiarity with computers can also limit the validity of CBT assessment scores for ELs (Choi, Kim, & Boo, 2003; Sawaki, 2001; Taylor, Kirsch, Jamieson, & Eignor, 1999). For instance, test takers who are not familiar with computers may fail to answer an item or even an entire section correctly, because they do not understand how to effectively and efficiently use the computer to complete the task. CBT continues to develop and may help to limit the challenges associated with making silent reading tests more valid, but, to date, they have had limited effectiveness in achieving this objective.

In a more recent look at the use of computers in silent reading assessment, Hiebert et al. (Chapter 9, this volume) conducted mode studies (i.e., studies that investigate the conditions of assessment administration) of the linkage between oral and silent reading rates and comprehension as measured by computer-administered testing and paper-and-pencil testing. Results indicated silent reading rates were faster than oral reading rates, and no differences were found between the testing conditions of paper-and-pencil versus computer-administered rate and comprehension tests. They also found that reading comprehension was not compromised by allowing students in fourth grade to read silently with accountability a comprehension test for the students in the top three quartiles of reading achievement. However, for those students scoring in the bottom quartile of reading achievement, oral reading with feedback was recommended.

Using Testing Accommodations to Aid in Assessing Silent Reading

With the aim of increasing the validity of assessment tasks for ELs, some research-ers and practitioners have turned to the use of accommodations, analogous to ones used by students with disabilities. In fact, a number of laws aimed at ensur-ing the inclusion of ELs in high-stakes assessments require the use of accommo-dations when assessing this population of students (Koenig & Bachman, 2004). Testing accommodations for ELs refer to changes in the testing process that help students demonstrate their actual abilities on the construct of interest despite their limited language proficiency—without providing them an unfair advantage over other students. The aim of such accommodations is to help students best demon-strate their actual reading abilities on reading tests (Afflerbach, 2007). A number of accommodations have been used when assessing the silent reading ability of ELs, including bilingual instructions and questions, linguistic modification, test preparation, extra time, oral instructions, and a glossary of key terms or bilingual dictionaries.

Two of the more popular and appealing accommodations are bilingual and linguistic modifications. Bilingual accommodations provide ELs with the instruc-tions and test questions in both English and the first language of each student who takes the test. Studies designed to assess this accommodation have compared the silent reading test scores (usually MC probed recall) of ELs given the accommoda-tion and ELs not given the accommodation. Although such studies have used large samples to maximize their power, the general finding has been a failure to find a difference between the scores of ELs who received instructions and comprehen-sion questions in both English and their first language and those who received them only in English (Anderson, Liu, Swierzbin, Thurlow, & Bielinksi, 2000).

Another accommodation, which has received a great deal of attention for im-proving the validity of assessments for ELs, is linguistic modification or simpli-fication. This accommodation has been used in reading assessments to limit the effects of assessment language on student test performance. Questions designed to assess students' reading comprehension are linguistically simplified, so the lan-guage in them does not create an additional challenge for the students. For in-stance, passive voice is changed to active, conditionals are replaced with separate phrases, and relative clauses are removed (Abedi & Lord, 2001). However, despite the attractiveness of this accommodation, little evidence has been garnered that suggests that it increases the validity of the silent reading scores of ELs (Francis, Rivera, Lesaux, Kieffer, & Rivera, 2006).

Research on the effects of other accommodations to assess the abilities of ELs has similarly failed to indicate that they effectively improve the validity of the assessments. A meta-analysis conducted by Francis et al. (2006) indicated that of seven of the more common accommodations used, only English-language dic-tionaries and glossaries were shown to significantly influence ELs' scores, and the effect size for this accommodation was very small. Thus, although the use of ac-commodations for assessing the reading abilities of ELs is popular and appealing

and current laws even require the approach, so far, they have had limited effectiveness in increasing the validity of ELs' scores on reading assessments.

Summing Up Problems and Perils in Assessing ELs' Silent Reading

It is clear that free and probed recall silent reading assessments such as retelling what was read or asking probing questions about what was understood are lacking. MC, matching, summary writing, story retell, and other similar task types all require skills not necessary for comprehending a text read silently and often change the typical processes students use while reading silently. Consequently, these methods introduce construct irrelevant variance into the silent reading scores that they yield and, hence, the obtained scores may not be considered representative indicators of students' silent reading ability. Future developments of CBT and testing accommodations may play a positive role in improving these approaches, but, to date, neither has been shown to markedly increase the validity of silent reading assessment scores.

The logic associated with question answering and retelling is that if students have read the text and comprehended it, they will be able to give back what the text was about or answer questions about it. Although this logic is both appealing and popular, it fails to shed light on the actual cognitive processing of L1s or ELs while reading silently. In silent reading, retelling or answering questions focuses on the "product" of the reading, not the process. Furthermore, retelling or question answering falls short of adequately measuring current conceptualizations of reading comprehension, such as those described in the theoretical work of Kintsch's (2004) construction-integration (CI) theory. To gain access to the unseen cognitive processes associated with silent reading, one must be able to pierce the curtain obstructing the view of students' silent reading processes and link these to product measures of silent reading. What might such assessments entail in the future?

Promising Directions for Silent Reading Assessment of ELs

The use of high-speed, infrared, eye-movement photography has long represented the hopes of those who would measure cognitive processing during silent reading (see Chapter 2, this volume). Eye movements known as saccades and fixations are photographed using high-speed, infrared tracking of the pupil of the eye as it moves along a line of print. The mapping of saccades and fixations onto the line of the print by the use of computer software programs allow measurement of silent reading processes, because past basic research has shown clearly that the effectiveness and efficiency of eye movements (i.e., saccades and fixations) along a

line of print are driven by cognitive processing factors (Rayner & Pollastek, 1989). Saccades, or short, jerking jumps of the eyes over print, indicate that the eyes are actively taking in print stimuli to be processed in the mind. Fixations, or points in the line of print where the eyes stop for a moment, are indicators that the eyes have taken in a chunk of printed language for processing in the mind and those visual stimuli are currently being processed.

The complexity of the eye-movement equipment programming, the need to bring students into a laboratory environment, and the cost of the necessary cameras and computer hardware were in the past prohibitive. Emerging technologies such as fMRIs, MRIs, and optical scanning that could be used to effectively link brain processing during silent reading suffer from the same problems as those associated with past generations of eye-movement photography—complex operations, cost, and nonportability.

However, contemporary eye-movement photography systems such as those produced by *Reading Plus* have become much less complex to operate, are highly portable, and are far less expensive. When using older eye-movement systems, students had to place their chins on stands and hold their heads very still for the cameras to operate properly. In next generation eye-movement photography systems, students wore helmets with pencil-sized cameras or had to remain confined to a small range of head movements while viewing text on a television screen, and their eyes were tracked with a fixed pan tilt camera. Today's eye-movement photography allows a student to wear a lightweight set of clear plastic goggles (much like safety glass goggles) to read a text held in their hands. Nevertheless, even with these advances current eye-movement photography has not advanced to the point of allowing researchers to measure reading of self-selected books. Instead, students must read texts that the eye-movement photography equipment developers have preselected and calibrated for assessment purposes. Although less ecologically valid than measuring students' silent reading eye movements of self-selected texts, the newer generation eye-movement photography equipment does allow for tracking eye movements during reading and, consequently, the cognitive processes ELs (and L1s) might be employing while reading silently.

Another interesting measure associated with eye movements is the duration of gaze, or how long the eyes take to make the short, jerky jump over a segment of print, usually a single word (Just & Carpenter, 1987). The duration of gaze measure has been hypothesized to be a measurement of how difficult a word is to process for either decoding or comprehension reasons. Some words are more difficult to decode because of length or structure. Other words take longer to process, because the meaning of the word is unfamiliar or difficult to retrieve from memory. Thus, the duration of gaze measurement may hold some promise in determining not only if one is processing or can process a word, but also how quickly and easily a word's visual, phonological, and lexical elements can be accessed. Also eye-movement photography may at some point provide an in-process means for determining which word meanings or vocabulary within a text obstruct reading fluency and comprehension. Eye-movement photography may also provide a

window into silent reading rates, both those in short text excerpts and those used in longer readings of connected text. Eye-movement photography could also one day provide insights on differing levels of students' silent reading stamina and how reading stamina as measured by efficient and effective eye movements may be associated with later reading acquisition and motivation.

The work of Hiebert et al. (Chapter 9, this volume) provides another promising avenue for assessing ELs' silent reading rates and comprehension. Assuming that reading comprehension tests could be devised for and norm referenced to an EL population, using the computer-based silent reading and comprehension testing format described by these researchers could provide classroom teachers with an effective and efficient process for measuring ELs' silent reading rates and comprehension in the future.

Another variable that obstructs, or in some ways at least obfuscates, the measurement of ELs' (and L1s') reading acquisition and engagement during silent reading is the invisible social context in which silent reading is often practiced— isolation or independence. Downing and Leong (1982) observe that most of our reading is done for our own private purposes and not in overtly observable social contexts. Going off and silently reading alone was presumed to be best practice for many years for ELs to acquire English reading facility and motivation (Freeman & Freeman, 2008; Krashen, 2002). In fact, in typically implemented SSR programs, asking students to talk about, discuss, or report in any way on their reading was seen as undesirable, having some of the alleged effects of the much-maligned written or oral book reports.

Going off alone to read impedes the beneficial human interactions around text that have been shown repeatedly to facilitate reading comprehension and motivate students to engage in sustained reading of texts (NICHD, 2000; Stahl, 2004). However, insofar as measuring ELs (and L1s') silent reading, the independence factor removes from view the overtly observable interactions around text that might provide a glimpse into students' comprehension processes and motivation. Even if such social exchanges were to be standard classroom practice during silent reading, this approach to measurement of silent reading suffers from the "product not process" measurement issues previously discussed in this chapter. However, it is clear that the private nature of independent reading provides yet one more obstruction to a clear view of silent reading processes and how these operate to influence reading achievement and motivation.

Advances in silent reading comprehension assessment are also beginning to emerge out of a general dissatisfaction expressed with the current comprehension assessment tools and processes (Paris & Stahl, 2005). One such example is Duke's (n.d.) Concepts of Comprehension Assessment (COCA). The COCA is designed to measure four contributors to reading comprehension: comprehension strategy use, vocabulary strategy use and knowledge, knowledge of informational text features, and comprehension of graphics in the context of text. The COCA was designed for use by classroom teachers, reading specialists, and paraprofessionals to inform their comprehension instruction and decision making, and it can be used by

researchers to evaluate students' silent or oral reading comprehension. The COCA represents new thinking about how to measure the multidimensional nature of reading comprehension that is aligned with current theories such as Kintsch's (2004) CI theory (Duke, 2005).

Conclusions About Assessing ELs' Silent Reading

There are many obstacles impeding the measurement of ELs' (and L1s') English print literacy acquisition and motivation during silent reading. Findings of the National Literacy Panel on Language-Minority Children and Youth (August & Shanahan, 2006) present an EL literacy acquisition empirical research base that is weak in both quality and quantity. There is even less known about how to effectively assess ELs' English literacy acquisition during silent reading.

For many years, silent reading was assumed to be one of the most effective practices for promoting ELs' English print literacy (Freeman & Freeman, 2008). The NRP (NICHD, 2000) found the existing body of evidence for independent, silent reading practice to be lacking in quality and quantity as well as failing to converge on a consistent finding that independent, silent reading practice yielded better reading fluency, achievement, or motivation results. The sudden cessation of independent, silent reading practice in classrooms also brought to a standstill attempts to understand how to assess independent, silent reading. Although independent, silent reading is now making a comeback in many classrooms, effective assessment of silent reading processes and products remains elusive. Silent reading is a cognitive activity not easily accessed through behavioral observations. The invisible social context in which silent reading is often practiced, quiet isolation, greatly frustrates and complicates researchers' attempts to assess silent reading processes. Past attempts at measuring ELs' silent reading process and products have centered almost exclusively on one of several less than satisfactory assessment approaches discussed in this chapter.

New technologies such as high-speed infrared eye-movement photography and brain function measures such as MRIs and fMRIs promise to provide new and exciting insights on the processes used in silent reading. This will be especially true when eye-movement photography can be directly linked to measures of cognitive processing, such as those now available through fMRI and other time-elapsed brain functioning measures. As researchers gain access to the otherwise hidden operations of silent reading processes, they will then be able to fashion increasingly sensitive, valid, authentic, and responsive silent reading assessment tools, protocols, and procedures. Although the current status of silent reading process and product assessment is frustratingly inadequate (Paris & Stahl, 2005), technological advances in measuring human information processing and newer multidimensional comprehension assessment tools hold out considerable promise for reading researchers to come to better understand and assess the silent reading of ELs.

QUESTIONS FOR PROFESSIONAL DEVELOPMENT

1. How are ELs different from L1s?

2. Describe the four methods discussed for assessing ELs' silent reading.

3. What are the problems and perils of each of these four methods?

4. How might these methods be adapted to limit these problems?

5. What is infrared rapid eye movement technology and what promises does it hold for assessing ELs' silent reading?

REFERENCES

Abedi, J., & Lord, C. (2001). The language factor in mathematics tests. *Applied Measurement in Education, 14*(3), 219–234. doi:10.1207/S15324818AME1403_2

Afflerbach, P. (2007). *Understanding and using reading assessment, K–12.* Newark, DE: International Reading Association.

Alderson, J.C. (2000). *Assessing reading.* New York: Cambridge University Press.

Alderson, J.C., & Huhta, A. (2005). The development of a suite of computer-based diagnostic tests based on the Common European Framework. *Language Testing, 22*(3), 301–320. doi:10.1191/0265532205lt310oa

Allington, R.L. (Ed.). (2002). *Big brother and the national reading curriculum: How ideology trumped evidence.* Portsmouth, NH: Heinemann.

Anderson, M., Liu, K., Swierzbin, B., Thurlow, M., & Bielinski, J. (2000). *Bilingual accommodations for limited English proficient students on statewide reading tests: Phase 2* (Minnesota Rep. No. 31). Minneapolis, MN: National Center on Educational Outcomes, University of Minnesota. Retrieved April 29, 2010, from www.cehd.umn.edu/NCEO/OnlinePubs/archive/AssessmentSeries/MnReport31.html

Au, K.H. (2009). Culturally responsive instruction: Application to multiethnic, multilingual classrooms. In L. Helman (Ed.), *Literacy development with English learners: Research-based instruction in grades K–6* (pp. 18–39). New York: Guilford.

August, D., & Shanahan, T. (Eds.). (2006). *Developing literacy in second-language learners: Report of the National Literacy Panel on Language-Minority Children and Youth.* Mahwah, NJ: Erlbaum.

Bachman, L.F., & Palmer, A.S. (1996). *Language testing in practice.* New York: Oxford University Press.

Barnwell, D.P. (1996). *A history of foreign language testing in the United States: From its beginnings to the present.* Tempe, AZ: Bilingual.

Brisbois, J.E. (1995). Connections between first- and second-language reading. *Journal of Reading Behavior, 27*(4), 565–584.

Butler, F.A., & Stevens, R. (2001). Standardized assessment of the content knowledge of English language learners K–12: Current trends and old dilemmas. *Language Testing, 18*(4), 409–427.

Choi, I., Kim, K.S., & Boo, J. (2003). Comparability of a paper-based language test and a computer-based language test. *Language Testing, 20*(3), 295–320. doi:10.1191/0265532203lt258oa

Downing, J.A., & Leong, C.K. (1982). *Psychology of reading.* New York: Macmillan.

Duke, N.K. (2005). Comprehension of what for what: Comprehension as a nonunitary construct. In S.G. Paris & S.A. Stahl (Eds.), *Children's reading comprehension and assessment* (pp. 93–104). Mahwah, NJ: Erlbaum.

Duke, N.K. (n.d.). *Informational comprehension assessment in the primary grades: Concepts of Comprehension Assessment (COCA).* East Lansing, MI: Literacy Achievement Research Center. Retrieved April 29, 2010, from www.msularc.org/html/project_COCA_materials.html

Elley, W.B. (2000). The potential of book floods for raising literacy levels. *International*

Review of Education, 46(3/4), 233–255. doi:10.1023/A:1004086618679

Elley, W.B., & Mangubhai, F. (1983). The impact of reading on second language learning. Reading Research Quarterly, 19(1), 53–67. doi:10.2307/747337

Fitzgerald, J., Amendum, S.J., & Guthrie, K.M. (2008). Young Latino students' English-reading growth in all-English classrooms. Journal of Literacy Research, 40(1), 59–94. doi:10.1080/10862960802070459

Francis, D.J., Rivera, M., Lesaux, N., Kieffer, M., & Rivera, H. (2006). Practical guidelines for the education of English language learners: Research-based recommendations for the use of accommodations in large-scale assessments. Portsmouth, NH: Center on Instruction, RMC Research Corporation. Retrieved September 9, 2009, from www.centeroninstruction.org/files/ELL3-Assessments.pdf

Freeman, Y.S., & Freeman, D.E. (2008). Academic language for English language learners and struggling readers: How to help students succeed across content areas. Portsmouth, NH: Heinemann.

Garcia, G.E., & DeNicolo, C.P. (2009). Making informed decisions about the language and literacy assessment of English language learners. In L. Helman (Ed.), Literacy development with English learners: Research-based instruction in grades K–6 (pp. 64–86). New York: Guilford.

Goodman, K.S., & Gollasch, F.V. (1980). Word omissions: Deliberate and non-deliberate. Reading Research Quarterly, 16(1), 6–31. doi:10.2307/747347

Goodman, Y.M., & Goodman, K.S. (1994). To err is human: Learning about language processes by analyzing miscues. In R.B. Ruddell, M.R. Ruddell, & H. Singer (Eds.), Theoretical models and processes of reading (4th ed., pp. 104–123). Newark, DE: International Reading Association.

Green, B.F. (1983). The promise of tailored tests. In H. Wainer & S. Messick (Eds.), Principals of modern psychological measurement (pp. 69–80). Hillsdale, NJ: Erlbaum.

Guthrie, J.T., & Wigfield, A. (1997). Reading engagement: Motivating readers through integrated instruction. Newark, DE: International Reading Association.

Helman, L. (Ed.). (2009). Literacy development with English learners: Research-based instruction in grades K–6. New York: Guilford.

Herrell, A., & Jordan, M. (2004). Fifty strategies for teaching English language learners (2nd ed.). Upper Saddle River, NJ: Prentice Hall.

Imao, Y. (2008). Investigating the effects of reading and writing on summaries written by L2 writers. Unpublished doctoral dissertation, University of California, Los Angeles.

Just, M.A., & Carpenter, P.A. (1987). The psychology of reading and language comprehension. Boston: Allyn & Bacon.

Kamil, M.L. (2008). How to get recreational reading to increase reading achievement. In Y. Kim et al. (Eds.), 57th yearbook of the National Reading Conference (pp. 31–40). Oak Creek, WI: National Reading Conference.

Kelley, M., & Clausen-Grace, N. (2006). R5: The sustained silent reading makeover that transformed readers. The Reading Teacher, 60(2), 148–156. doi:10.1598/RT.60.2.5

Kelly, F.J. (1915). The Kansas silent reading test. Topeka: Kansas State Printing Plant.

Kindler, A.L. (2002). Survey of the states' limited English proficient students and available educational programs and services: 2000–2001 summary report. Washington, DC: National Clearinghouse for English Language Acquisition & Language Instruction Educational Programs.

Kintsch, W. (2004). The construction–integration model of text comprehension and its implications for instruction. In R.B. Ruddell & N.J. Unrau (Eds.), Theoretical models and processes of reading (5th ed., pp. 1270–1328). Newark, DE: International Reading Association.

Koenig, J.A., & Bachman, L.F. (Eds.). (2004). Keeping score for all: The effects of inclusion and accommodation policies on large-scale educational assessments. Washington, DC: National Academies Press.

Krashen, S. (2002). More smoke and mirrors: A critique of the National Reading Panel report on fluency. In R.L. Allington (Ed.), Big brother and the national reading curriculum: How ideology trumped evidence (pp. 112–124). Portsmouth, NH: Heinemann.

Martínez-Roldán, C.M., & López-Robertson, J.M. (1999). Initiating literature circles in a first-grade bilingual classroom. The Reading Teacher, 53(4), 270–281.

Mohr, K.A.J. (2003). Children's choices: A comparison of book preferences between Hispanic and non-Hispanic first-graders. Reading Psychology, 24(2), 163–176. doi:10.1080/02702710308231

Montero, M.K., & Kuhn, M.R. (2009). English language learners and fluency development: More than speed and accuracy. In L. Helman (Ed.), *Literacy development with English learners: Research-based instruction in grades K–6* (pp. 156–177). New York: Guilford.

National Center for Education Statistics. (2004). *The condition of education 2004* (NCES 2004-077). Washington, DC: U.S. Department of Education. Retrieved April 29, 2010, from nces.ed.gov/pubs2004/2004077.pdf

National Institute of Child Health and Human Development. (2000). *Report of the National Reading Panel. Teaching children to read: An evidence-based assessment of the scientific research literature on reading and its implications for reading instruction* (NIH Publication No. 00-4769). Washington, DC: U.S. Government Printing Office.

Ockey, G.J. (2009). Developments and challenges in the use of computer-based testing for assessing second language ability. *The Modern Language Journal, 93(Special Focus Issue)*, 836–847. doi:10.1111/j.1540-4781.2009.00976.x

Paris, S.G., & Stahl, S.A. (Eds.). (2005). *Children's reading comprehension and assessment.* Mahwah, NJ: Erlbaum.

Pilgreen, J.L. (2000). *The SSR handbook: How to organize and manage a sustained silent reading program.* Portsmouth, NH: Boynton/Cook.

Proctor, C.P., Carlo, M., August, D., & Snow, C. (2005). Native Spanish-speaking children reading in English: Toward a model of comprehension. *Journal of Educational Psychology, 97(2)*, 246–256. doi:10.1037/0022-0663.97.2.246

Rayner, K., & Pollatsek, A. (1989). *The psychology of reading.* Hillsdale, NJ: Erlbaum.

Reutzel, D.R., Jones, C.D., Fawson, P.C., & Smith, J.A. (2008). Scaffolded silent reading: A complement to guided repeated oral reading that works! *The Reading Teacher, 62(3)*, 194–207. doi:10.1598/RT.62.3.2

Ross, S. (1998). Self-assessment in second language testing: A meta-analysis and analysis of experiential factors. *Language Testing, 15(1)*, 1–20.

Sawaki, Y. (2001). Comparability of conventional and computerized tests of reading in a second language. *Language Learning & Technology, 5(2)*, 38–59. Retrieved February 12, 2008, from llt.msu.edu/vol5num2/sawaki/default.html

Stahl, S.A. (2004). What do we know about fluency? Findings of the National Reading Panel. In P. McCardle & V. Chhabra (Eds.), *The voice of evidence in reading research* (pp. 187–211). Baltimore: Paul H. Brookes.

Stahl, S.A., & Heubach, K. (2006). Fluency-oriented reading instruction. In K.A.D. Stahl & M.C. McKenna (Eds.), *Reading research at work: Foundations of effective practice* (pp. 177–204). New York: Guilford.

Taillefer, G.F. (1996). L2 reading ability: Further insight into the short-circuit hypothesis. *The Modern Language Journal, 80(4)*, 461–477.

Taylor, C., Kirsch, I., Jamieson, J., & Eignor, D. (1999). Examining the relationship between computer familiarity and performance on computer-based language tasks. *Language Learning, 49(2)*, 219–274. doi:10.1111/0023-8333.00088

Wallace, C. (1992). *Reading.* New York: Oxford University Press.

Yamashita, J. (2002). Mutual compensation between L1 reading ability and L2 language proficiency in L2 reading comprehension. *Journal of Research in Reading, 25(1)*, 81–95. doi:10.1111/1467-9817.00160