

Running Head: FLUENCY AND ACCURACY

Variables that Affect the Correlation between Fluency and Accuracy
with a Measure of Reading Comprehension

Michelle K. Hosp

Peabody College of Vanderbilt University

Abstract

The purpose of this review was to examine the variables (e.g., grade, length of passage or word list, criterion measure) that affect the correlation between each of four different types of word reading measures (text fluency, list fluency, text accuracy and list accuracy) and a criterion measure of reading comprehension. Twenty-seven individual studies were identified, however, many reported correlations with more than one type of word reading measure. Due to the overlap of participants across the four measures of word reading, overall correlations across these four measures could not be compared to determine if the scores were significantly different from each other. Significance testing was, however, conducted on the four different types of word-reading measures revealing some variables do impact the magnitude of the correlation for some types of word-reading measures.

Variables that Affect the Correlation between Fluency and Accuracy with a Measure of Reading Comprehension

Reading is the most critical skill students will learn and one of the best predictors of overall success in school (Stanovich, 1986), and in society (Lyon, 1997). Students who struggle in reading have a higher rate of dropping out of school and hold lower-paying jobs (Snow, Burns, & Griffin, 1998). Additionally, illiterate adults account for 75% of the unemployed, 33% of mothers receiving aid, 85% of juveniles appearing in court, 60% of prison inmates, and 40% of minority youth (Orton Dyslexia Society, cited in Adams, 1990). Moreover, 2.8 million students ages 6 to 21 are identified as having a learning disability (U.S. Department of Education, 2000), representing the majority of students receiving special education services (51.0%) (U.S. Department of Education, 2000). Among those students identified as LD, 80% have serious problems in reading (Snow et al. 1998).

The ability to read words (i.e., to decode) is often used to make the distinction between individuals who are proficient in reading from individuals who struggle or have a reading disability (Adams, 1990; Shankweiler & Liberman, 1972; Stanovich, 1980; Stanovich, 1986). Therefore, research in reading disabilities has primarily focused on the remediation of word-reading problems (Jenkins, Fuchs, Espin, Broek, & Deno, 2000). In addition, word-reading skills are highly correlated with reading comprehension (Gough, Hoover, & Peterson, 1996).

Two indices commonly used to measure word-reading skills are fluency and accuracy. Fluency is the number of words read correctly in a fixed amount of time (usually indexed per minute); accuracy is the percentage of words read correctly. In addition, fluency and accuracy can be used to describe word reading when students read connected text or word lists. Throughout this paper we use the term text fluency to refer to the number of words read correctly per minute in text; text accuracy to refer to the percentage of

words read correctly in text; list fluency to refer to the number of words read correctly per minute on word lists; and list accuracy to refer to the percentage of words read correctly on word lists.

While word reading can be measured in these four different ways, many variables may affect the strength of the correlation between word-reading skills and reading comprehension. For example, when text or word lists are easy or below a student's grade level, a ceiling effect can occur, impacting the correlation. Moreover, Jenkins and Jewell (1993) reported that students in lower grades (second through fourth) had a higher correlation between text fluency and comprehension than did students in higher grades (fifth and sixth). While numerous studies have reported correlations between word-reading skills and comprehension, there has been no synthesis of these studies specifically looking at the variables affecting the magnitude of the correlation. Therefore, this is the first review to examine variables affecting the relation between word reading, as measured by fluency or accuracy in text or on word lists, with a criterion measure of comprehension. This introduction begins by providing an empirical rationale for focusing on word-reading skills as an indicator of comprehension. Then, reading theories that support word-reading as an indicator of overall reading ability are discussed. Finally, the importance of this synthesis in relation to research and practice in the classroom is provided.

Empirical Rationale for Focusing on Word Reading

Ehri and Saltmarsh (1995) demonstrated the importance of word reading when they found beginning first-grade readers were more proficient at word reading than older students with reading disabilities. It stands to reason that a failure to read words negatively impacts text comprehension (Snow, Burns, & Griffin, 1998). Moreover, most children with reading disabilities demonstrate satisfactory listening comprehension, which can then be utilized once word-reading problems are remediated (Jenkins et al., 2000). This is further supported by research demonstrating that word-reading skills, along with listening comprehension, account for virtually

all of the variance in reading skills (Hoover & Gough, 1990). Even when listening comprehension is controlled for, word reading has been shown to account for a large amount of variance in the difference between individuals' reading performance (Curtis, 1980).

In addition, teachers tend to be more concerned with word reading accuracy than reading fluency (Nathan & Stanovich, 1991). For example, Fuchs, Fuchs, Hosp, and Jenkins (in press) found that informal reading inventories (e.g., Woods & Moe, 1985) directed teachers to use accuracy rather than fluency when placing students in reading material. Moreover, the importance of word reading is further supported by theories focusing on some version of limited-capacity or verbal efficiency theory, which purports that word reading is at the center of reading comprehension (e.g., LaBerge & Samuels, 1974; Posner & Snyder, 1975; Perfetti, 1985; Stanovich, 1980). Three such theories are discussed below.

Theories of Reading that Support the Relation Between Word Reading and Comprehension

LaBerge and Samuels' theory of automaticity. This theory provides support that reading words fluently must occur before comprehension is possible. It is based on the following premise: "It is assumed that we can only attend to one thing at a time, but we may be able to process many things at a time so long as no more than one requires attention" (LaBerge & Samuels, 1974, p. 295). LaBerge and Samuels (1974) proposed that reading involves the coordination of several component processes that occur simultaneously.

According to Samuels (1987), this theory is based on four key elements: attention, visual memory, phonological memory and semantic memory. He defined attention as the "effort or energy required to perform cognitive tasks" (Samuels, 1987 p. 18). Five different types of attention are discussed relative to this theory. The first type of attention is overt, which is observable and makes it possible to predict if a person is paying attention or not. Second is covert, which equates to the level of arousal (i.e., alertness) a person is in when learning. The third type of attention is vigilance, meaning being able to focus on the task at hand for long periods of

time. Fourth, selective attention refers to the ability to block out unwanted distracters so as not to disrupt attention away from the task at hand. Last is capacity, which, unlike the first four that refer to types of attention, refers to the fixed amount of attention that is available.

The second key element to LaBerge and Samuels' theory is visual memory. This relates to the ability to extract the printed letters from the page to form words. Individual letters are the smallest unit; letter combinations are the next smallest unit, and whole words are the largest unit. A novice reader may focus on the individual letters to form words while the more experienced reader may focus on the entire word. Related to this is the third key element of LaBerge and Samuels' theory, phonological memory. Phonological memory refers to the sound units (i.e., phonemes or morphemes) that map onto the letters of words. Not only must individuals be able to see the letters or words but, they must also know what sounds go with what letters and words. Last, semantic memory refers to the declarative and procedural knowledge that is stored in relation to words and their meanings.

LaBerge and Samuels' (1974) theory is based on two notions. First, an individual has only so much accessible attention and attention cannot be divided among tasks. Second, practicing a skill allows it to become automatic and therefore require little to no attention (Slocum, Street, & Gilberts, 1995). When skills become automatic, this allows for the ability to perform multiple tasks simultaneously (Samuels & Flor, 1997). Hence encoding letters must be automatic to read words and reading words must be automatic in order for attention to be used for comprehension. If identifying letters and words is not accurate and automatic, then attention will be spent on identifying letters and words and not on comprehending what is read. According to Samuels and Flor (1997), performing multiple tasks concurrently such as decoding and comprehension indicates that decoding is automatic.

Moreover, Samuels and Flor (1997) proposed that a skill must be performed accurately before it can become automatic and therefore require little to no attention. An individual may be accurate and still require a lot of attention. It is only when the skill is

over-learned that it no longer requires attention. This point is further illustrated by looking at two students in the same class where one reads 100 words correct in one minute with no errors and the other reads 30 words correct in one minute with no errors. Although they are both accurate, one student is more automatic as implied by the number of words read. According to this theory, automaticity at the word level allows for attention to be placed on reading comprehension. Notice that automaticity of comprehension is not included in this theory. Instead, only those tasks that are predictably the same (e.g., word recognition) and tasks that are not always changing (e.g., text comprehension) are considered appropriate to become automatic (Samuels & Flor, 1997). LaBerge and Samuels' (1974) theory of automaticity relies on a bottom-up approach to reading where word reading must first occur before comprehension is possible. Problematic with bottom-up theories is that they do not address what happens when word reading is slow and inefficient. Posner and Snyder's (1975a, 1975b) two-process theory of contextual expectancy provides some guidance related to the problem of word reading that is slow and inefficient.

Posner and Snyder's two-process theory of contextual expectancy. This theory of reading expands on the one proposed by LaBerge and Samuels (1974) by using two contextual mechanisms. The first is the automatic activation of specific pathways in the nervous system, indicating that items sharing the same pathways are processed more quickly. In addition, pathways that are confined to the memory system will not inhibit the processing of items whose pathways are not activated (Posner & Snyder, 1975a, 1975b). For example, when the word PIG is activated in the memory system it has no effect on the activation of the word HOT. While activating the word PIG assists the processing of signals related to it, there is no inhibitory effect of the activation. Although this first type of contextual mechanism requires time, it does not require conscious attention. This is the difference between the first and the second type of contextual mechanism Posner and Snyder (1975a, 1975b) use in their two-process theory.

The second contextual mechanism occurs when conscious attention is given to the information (i.e., word). This increases the benefit from the pathway that is activated, however it inhibits other signals from being activated. Because attention is being used to understand the information, this interferes with other signals that may be related and could have been automatically activated. These two types of contextual mechanisms provide support that context can be used to help aid word reading although the more context is used, the less conscious attention is available for comprehending what is read. While LaBerge and Samuels (1974) and Posner and Snyder (1975a, 1975b) both claim that efficient low-level word-reading skills free up cognitive processes that can be used for comprehending text, it is Stanovich (1980) who further explains which type of reader (i.e., poor or good) relies on context to help aid word reading.

Stanovich's interactive-compensatory model. Stanovich (1980) proposed a theory of reading, building on Posner and Snyder's two-process theory of contextual expectancy, by expanding on the individual differences in context reading fluency. According to his model, context is used to facilitate word recognition for both good and poor readers, although for different reasons. Whereas the top-down model suggests only good readers use context to help read words, Stanovich (1980) proposed that poor readers may actually use context more to aid word reading. This occurs because they do not have good decoding skills; therefore, they rely on the context to help them decipher unfamiliar words. On the other hand, good readers may also use context, but they may do so to monitor comprehension rather than reading words. For example, good readers have the ability to process printed words rapidly and efficiently so they do not have to rely on context, but instead can use their mental energy to comprehend what they read. The only time they would rely on context is to read a word that is unfamiliar to them.

In addition to context facilitation, Stanovich (1980) proposed that context-free word recognition (e.g., lists) at the word and subword level also provides important information on individual differences in reading fluency. He described three distinct abilities

related to context-free word recognition. First, the ability to recognize words automatically allows mental processing to focus on higher-level skills (comprehension). Second, the rate at which words are recognized and coded into short-term memory aids the integrative comprehension processes that helps activate the information that is already there (Lesgold & Perfetti, 1978). Last, phonological abilities aid word reading by providing a predictable pathway for letters to be accessed and also by providing a predictable code for information held in short-term memory.

Importance of Synthesis for Research and Practice

Although these theories support word reading as necessary for reading comprehension, they do not specifically address the variables that may affect the magnitude of the correlation between measures of word-reading skills and comprehension. This is important because students with reading disabilities often have deficits in their word reading ability. And word-reading skills may be differentially correlated with comprehension based on certain variables (e.g., students' grade, difficulty of material read, whether the student has a disability). Moreover, teachers and researchers commonly use word-reading skills as an indication of performance or to judge the effectiveness of reading interventions (Jenkins et al., 2000). While word-reading skills have been shown to be highly correlated with text comprehension (Gough, et al. 1996), the variables that affect this correlation have not been examined in relation to fluency and accuracy in text and on word lists. Therefore, this review examines the literature on the relation between word-reading skills and comprehension and the effect specific variables have on these correlations when using fluency and accuracy as a measure of word reading in text and on word lists.

Method

Literature Search

An exhaustive literature search was conducted to identify studies reporting correlations between a measure of reading comprehension and performance on four ways of indexing word- reading skill: text fluency, text accuracy, list fluency, and list accuracy. First, a computerized search using Dissertation-Abstracts, ERIC, Exceptional Child Educational Resources, Education Abstracts and PsycINFO was conducted using the following descriptors: reading accuracy, reading fluency, reading rate, reading speed, assessment, automaticity, fluency, reading, and word decoding. Each descriptor was entered in isolation so that the largest number of studies would be identified. From there, the descriptor reading comprehension was entered to produce a manageable field of citations to search.

Second, a manual search was conducted from 1980 to 1999 in the following journals: Exceptional Children, Journal of Learning Disabilities, Journal of Reading Behavior, Journal of School Psychology, Learning Disability Quarterly, Reading Research Quarterly, Remedial and Special Education, School Psychology Review, and The Journal of Special Education. Each journal was chosen because of its likelihood to publish studies related to word reading. Third, the reference sections from collected studies were searched to determine additional citations.

This search yielded 1,407 relevant articles. Of these studies, 21 provided correlations for text fluency; 7 for list fluency, 3 for text accuracy, and 7 for list accuracy. However, these studies were not mutually exclusive. For example, a study may have addressed both text fluency and list fluency; such a study would be counted in each area. A total of 27 individual studies were used (see Table 1).

Criteria for Study Inclusion

For studies to be included, they had to meet the following criteria: (a) participants must have been in first through eighth grade; (b) the study had to use real words to obtain scores on the word reading measures; (c) word reading had to be conducted orally; (d) a measure of reading comprehension had to be used as a criterion measure; (e) correlations had to be provided between the word reading measure and the criterion measure; (f) and the study must have been conducted in English.

The following list provides a sample of studies that were excluded from the review along with the reasons for their exclusion: Armstrong (1983) used a single subject design; Lopez, Clark, and Winer (1979) used silent reading to calculate words read correctly per minute; Aaron, Joshi, and Williams (1999); DeSoto and DeSoto (1983); Rasinski (1990); and Wolf, Bally, and Morris (1986) all used the number of seconds to read a list or passage instead of words read correctly per minute; Biemiller (1977-1978) and Roberts and Smith (1980) did not provide correlations between fluency and a criterion measure; Perfetti and Hogaboam (1975) did not provide correlations for list accuracy and a criterion measure; Stanovich, Cunningham, and West (1981) used a rapid naming task; and Stanovich (1981) and Perfetti, Finger, and Hogaboam (1978) both did not use a criterion measure.

Methods for Analyzing Studies

To draw comparisons within and across the four types of word-reading measures (text fluency, text accuracy, list fluency, list accuracy), studies were coded and mean correlations were calculated. The studies were coded based on how word reading was measured (i.e., text fluency, list fluency, text accuracy, and list accuracy), number of participants, student group (i.e., general education, special education, at-risk), grade, type of passages (i.e., basal reader, researcher written) or type of word list (i.e., basal reader, researcher developed, commercial test), number of words per passage or word lists, grade level of passages or word lists, type

of criterion measure (i.e., mazes, standardized, group-administered test, basal reader test, questions, story retell with questions, or specific individually administered tests), along with correlations reported in each study.

This information was then entered into SPSS so that correlations could be averaged together, weighting for sample size. This provided mean correlations for specific variables as well as an overall correlation for each of the four types of word-reading measures. The variables examined for text fluency were type of passage (i.e., basal or researcher written); level of passage (i.e., below, at, or above grade level); length of passage; whether the passage was based on students' grade level or a common passage (i.e., a generic passage written to be used with students of all grade levels); criterion measure; grade; student group; and sample size. The variables examined for list fluency were type of word list (i.e., basal or vocabulary series, or from a passage); level of word list (i.e., below, at, or above grade level); length of word list; criterion measure; grade; student group; and sample size. The variables examined for text accuracy were type of passages (i.e., basal or researcher written); length of passage; criterion measure; grade; student group; and sample size. And the variables examined for list accuracy were type of word list (i.e., commercial tests, researcher developed); length of word list; criterion measure; grade; student group; and sample size. The same variables were not used for each of the four word reading measures because of the lack of information provided in the studies. For example, passage level and list level were not used to examine text fluency or list fluency because this information was not reported in the studies. In addition to obtaining mean correlations, significance testing was conducted between the correlations using the formula, $Z = (Z_1 - Z_2) / \{[(1/(n-3)) + (1/(n-3))]\}^{1/2}$. This provided information about whether differences between correlations were reliable.

In addition, each study was coded based on methodological qualities. Studies were coded on sample size, whether or not a clear description of the participants was provided (i.e., sex, student placement, setting), if authors provided a complete description of the materials and measures used and administration procedures, along with the length of passages or word lists used. The points awarded for sample size were: 3 (above 99), 2 (between 50 and 99), and 1 (below 50). Studies were also awarded 1 point each if they gave the sex of the students, placement of the students, setting, a description of the materials used, a description of the criterion measure, and how these measures were administered. No points were awarded if these descriptions were not provided. The points awarded for length of passage or list used were: 3 (passage over 250 words, list over 200 words), 2 (passage between 100 and 250 words, list between 100 and 200 words), 1 (passage or list less than 100 words). The possible range of points awarded was between 2 and 12. However, the actual range obtained was between 8 and 11, indicating they all did a reasonable job describing what occurred and with whom. Unlike intervention studies that can be coded for numerous threats to internal and external validity, these studies could only be coded on a limited number of variables because they were only providing correlations between word reading and a criterion measure of reading comprehension. Because methodological quality did not discriminate between studies, it was not included as a variable for analysis.

In addition, due to dependency in the data, not all correlations reported in the 27 studies were used. For example, if a study used two separate criterion measures with the same students, there would be dependency in the data because the same students are represented in both scores. This is important because in order to test the differences between correlations with dependent data,

information is required that is not typically reported in studies. However, some correlations computed on the same sample of students were averaged for some analyses.

There were times, however, when this was not appropriate. For example, if a study reported two or more correlations for different criterion measures with the same population, the averaged correlation could not be used because it would include data for more than one criterion measure. Moreover, the separate correlations could not be used because of dependency. Therefore, not all of the correlations reported in each section were used in every analysis.

Results and Discussion

How do Correlations Compare for the Four Types of Word Reading Measures?

For text fluency, 21 studies were identified with 33 correlations; the mean correlation was .69 (SD = .11). For list fluency, seven studies were identified, with seven correlations; the mean correlation was .60 (SD = .08). Text accuracy had five correlations that were derived from five studies, with a mean correlation of .62 (SD = .18). List accuracy had seven studies that provided seven correlations with a mean of .78 (SD = .12).

Due to the overlap of participants used across the four measures of word-reading skill, a large number of studies had dependency in their data. Therefore, the correlations from these studies could not be used to examine if the mean scores were significantly different from each other. Moreover, it was difficult to compare correlations for the four measures of word reading because of the difference in the number of studies reported in each area, and variability in sample size, grade, and criterion measures used. In addition, the mean correlations fluctuated depending on certain variables (e.g., sample size, criterion measures, length of

passages or word lists). Nevertheless, when possible, comparisons across specific variables for all four measures of word-reading skill were conducted, in addition to comparing the variables within each word reading measure.

Correlations for Text Fluency

Two types of passages were used to index text fluency. The majority of the correlations (72%) used passages from basal readers (i.e., Ginn; Holt, Rinehart, & Winston; MacMillan; Harcourt-Brace-Jovanovich; Allyn-Bacon; Scott-Foresman; or Houghton-Mifflin), with a mean correlation of .66 ($SD = .15$). The remaining 28% used passages developed for measuring student progress (i.e., Comprehensive Reading Assessment Battery; Reading Progress Monitoring Passages; or the Wheldall Assessment of Reading Passages), with a mean correlation of .74 ($SD = .10$). There was a significant difference between these two types of passages, $Z(1, 2341) = 3.87, p < .001$. One explanation for this difference may be related to the level of the passage used (i.e., below, at, or above grade level). The author-made passages were mostly below grade level, whereas the basal passages were mostly at grade level. This is explored further below along with differences in correlations based on the passage length, whether the passage was graded or common, criterion measure, grade, student group, and sample size.

Text fluency and difficulty level. As already noted, the difficulty of the passage (i.e., below, at, or above the students' grade level) may affect the size of the correlation between text fluency and comprehension. Shinn, Good, Knutson, Tilly, and Collins, (1992) suggested that studies using passages considered easy should be interpreted with caution. This is due to the possibility of students at upper grade levels reaching a ceiling, which would result in a nonnormal distribution. A ceiling effect occurs when the range of difficult items (or, in this case, the level of the passages) is restricted so that many of the students earn the maximum score. This may

have affected the passages because those below grade level would have been easier for students to read, meaning they may have read more words correctly than if the passages were at or above their grade level.

To determine if the difficulty of the passage was a factor, correlations for the studies using below, and at, grade level passages were compared. No correlations were reported for passages above grade level; therefore, this level was not used for the analysis. Only 20 of the 33 correlations could be used in this analysis because of dependency in the data, mixing grades, and using various difficulty levels within a grade. As shown in Table 2, the mean correlation for studies using below grade level passages (.73) was higher than those using passages at grade level (.68). This difference was significant, $Z(1, 1571) = 1.91, p < .05$. The reason for this difference may be due to the ceiling effect, as Shinn et al. (1992) had cautioned.

At the same time, some evidence suggests that passage level may not be a factor. Markell (1991) looked at the correlations for passages that were below, at, and above the students' grade level. While this study was not included in the mean correlations due to dependency in the data, it does offer a different perspective on level of passages used. Using the same 42 third-grade students, Markell provided correlations for each of the three passage levels (below, at, above grade) with two criterion measures (mazes and questions). What he found is in contrast to the information shown in Table 2. For maze, the correlations reported were .86, .89, and .87 for below, at, and above, grade level passages, respectively. This indicates no difference between correlations as a function of the level of text difficulty. However, when questions were used as the criterion measure, a different trend emerged. The correlations were .26, .50, and .40 for below, at, and above grade level passages. These correlations differ more than those shown in Table 2 or those described above for the maze criterion measure.

This difference may be related to the criterion measure, which may explain differences in the correlations (discussed below). Moreover, Markell did not find that easier passages increased the correlation with comprehension, as Shinn et al. (1992) had cautioned. Of course, another factor that should be considered is the length of passages used. This may provide further information explaining the magnitude of the correlation between text fluency and comprehension.

Text fluency and length. The length of passage used may have implications for the strength of the correlation between text fluency and comprehension. The reason for this is that students may complete a shorter passage before one minute, requiring examiners to prorate scores. Prorating is accomplished by multiplying the number of words read correct by 60 and dividing that product by the number of seconds actually read. This transformation could affect the size of the correlation by increasing the spread in scores, thus increasing the correlations. For example, if a student read 90 words correctly from a 100-word passage in 50 seconds, his/her prorated score would be 108. If the same student were given a 200-word passage, it is possible that they would read less than 108 words correctly per minute. Therefore, prorating makes assumptions about what the student may have done without actually measuring the full array of their skills.

Of the included studies, the length of the passages ranged from 90 to 400 words. The most common length was 250 words, used in 34% of the studies, while studies using passages of 90, 100, and 172 words were the least common, each accounting for only 3%. When a range of words per passage length was provided, the average was calculated based on the range. For example, if the authors reported that passages ranged from 132 to 213 words, these were averaged together to provide the mean length of passage

used. This occurred in two studies. The first one used passages with an average length of 172 words; the second, 190 words. In addition, one study did not report the number of words per passage.

Table 3 provides a breakout of the studies grouped by passage length. The highest mean correlation (Fuchs, 1981; see also Fuchs & Deno, 1992) occurred when passages of 100 words were used, and the authors indicated that scores were prorated when necessary. This mean correlation (.91), was significantly higher when compared to all other mean correlations: 90- word passages (.62), $Z(1, 155) = 4.87, p < .001$; 172-word passages (.63), $Z(1, 131) = 4.09, p < .001$; 190-word passages (.72), $Z(1, 424) = 5.17, p < .001$; 200-word passages (.70), $Z(1, 336) = 5.31, p < .001$; 250-word passages (.57), $Z(1, 774) = 7.77, p < .001$; 300-word passages (.73), $Z(1, 218) = 4.31, p < .001$; and 400-word passages (.83), $Z(1, 237) = 2.52, p < .01$.

Two other studies reported prorating and used passages with an average length of 172 words (Markell, 1991) and 200 words (Hintze, Shapiro, Conte, & Basile, 1997). Markell's correlations ranged from .26 using questions as the criterion measure to .89 using a maze task as the criterion. Based on this range, it is unclear if prorating affects the correlation. Hintze et al. also used a maze task, but reported correlations of .64 and .67, which is markedly different than the correlation reported by Markell. In addition, because other studies did not report whether students' scores had been prorated, it is difficult to determine what, if any, effect score transformation has on the correlation between text fluency and comprehension.

Evidence that further conflicts with prorating as a cause of inflating scores was observed by 400-word passages used by (Jenkins et al., 2000) and (Fuchs, Fuchs, & Maxwell, 1988), where prorating would not have been necessary. These 400-word passages provided the second highest mean correlation of (.83) which was significantly higher for all other passage lengths except, of

course, the 100-word passages. The difference between 90-word passages was (.62), $Z(1, 212) = 3.07, p < .01$; 172-word passages (.63), $Z(1, 188) = 2.48, p < .01$; 190-word passages (.72), $Z(1, 481) = 2.81, p < .01$; 200-word passages (.70), $Z(1, 393) = 5.90, p < .001$; 250-word passages (.57), $Z(1, 831) = 5.90, p < .001$; and 300-word passages (.73), $Z(1, 275) = 2.13, p < .05$. Given that 400-word passages also provided a large correlation along with the small number of studies reporting using prorating, no conclusions can be drawn concerning whether prorating affects the correlation between text fluency and comprehension.

Text fluency and graded or common passages. Another way to look at the passages is to determine if they were chosen based on students' grade level or if they were common. Common passages are those chosen for story content regardless of their readability level, whereas passages chosen specifically because the material was related to a students' grade level (below, at, or above) would be considered graded passages. For example, the passages from the Comprehensive Reading Assessment Battery and the Wheldall Assessment of Reading Passages are considered common passages because the authors developed them to be used with all students, regardless of the students' grade level.

Table 4 provides a breakdown of the correlations for second through sixth grade based on whether passages were graded or common. All of the graded passages were at the students' grade level, while the common passages could have been below, at, or above grade level. The only significant difference between common passages and passages that were specific to a student's grade level occurred at fourth grade. The mean correlation for the common passages (.76) was significantly different from the mean correlation for passages at the students' grade level (.56), $Z(1, 372) = 3.10, p < .001$. This difference may have occurred due to one study by Kranzler, Brownell, & Miller (1998) that reported an extremely low correlation of .41 between text fluency and the Kaufman Test of

Educational Achievement- Brief. Without this study, the mean correlation for passages at the students' grade level is .74, which is not significantly different from the common passage (.76), $Z(1, 315) = .28$, *ns*. Similar to written retell (discussed above) it is possible that the criterion measure had more impact on the correlation than the passage read.

Text fluency and criterion measure. Of the included studies, eight criterion measures were used. However, because of dependency in the data, only five of the eight criterion measures are discussed. The most common criterion measures (58%) were standardized, group-administered tests (i.e., California Achievement Test; Comprehensive Tests of Basic Skills; Diagnostic Reading Scales; Iowa Test of Basic Skills; Gates-MacGinite Reading Test; Metropolitan Achievement Test; Science Research Associates Achievement Series; Stanford Achievement Test; the Stanford Diagnostic Reading Test). The second most common criterion measures (19%) were maze passages. Fifteen percent used comprehension tests developed by the basal reading series (i.e., Scott-Foresman, Holt, Ginn; Houghton-Mifflin; Gates-MacGinite); while questions from the passages read and the Kaufman Test of Educational Achievement-Brief each made up 4% of the measures.

Table 5 shows the breakdown of the correlations by type of criterion measure used. The highest mean correlation for a criterion measure was reported for the maze (.78). This was significantly different from the standardized, group-administered tests (.69), $Z(1, 1656) = 2.66$, $p < .01$; and the Kaufman Test of Educational Achievement- Brief (.41), $Z(1, 267) = 3.00$, $p < .001$. A maze requires the student to read a passage with the first sentence intact. Thereafter, every *n*th word is deleted and replaced with three or five choices. Students read the passage and for each blank, choose one word that meaningfully restores the text. It is possible the maze criterion produced the highest correlation because the word-reading tasks and the criterion measure are markedly similar.

The second highest correlation (.71) was shared by two criterion measures (basal reader test and questions on the passage read). The basal reader test and questions on the passage both yielded significantly different correlations from those involving the Kaufman Test of Educational Achievement- Brief. The difference between the basal reader test and the Kaufman Test of Educational Achievement- Brief (.41) was, $Z(1, 195) = 2.81, p < .01$. The difference between the questions on the passage and the Kaufman Test of Educational Achievement- Brief was, $Z(1, 105) = 2.26, p < .05$.

While the mean correlation for basal readers' comprehension measures and questions on passages read were identical, the tasks involved differ greatly. For basal readers' comprehension measures, students performed two or more of the following tasks: identify synonyms for identified words, complete analogies by reading a story and selecting the similarity between two objects from a list, identify the main idea by reading a story and determining if a given sentence is the main idea or a supporting detail or neither, answer comprehension questions after reading a passage, answer true/false questions after reading a passage, and/or select the best meaning for a word that is underlined in the text. As for questions on passages, there are many different types of questions that can be posed. For example, students can be asked to recall details with who, what, where, when, and why questions, make predictions, or identify main ideas in a story. Another distinction is that question answering criterion measures may be derived from commercial tests (e.g., Neale Analysis of Reading- Revised comprehension subtest) or written by the researcher, which was the case for two of the three studies using questions on passages.

Given these differences, it is not surprising to see variability (range = .26 to .84) across the three studies that used questions. Markell (1991), who investigated the effects of increased oral reading fluency on two measures of reading comprehension, reported

consistently low correlations (e.g., .26, .50, and .40) for questions on passages. Fuchs et al. (1988), who investigated correlations with different criterion measures of reading performance, reported a high correlation of .84 with questions on passages. Additionally, the correlation derived with a commercial set of questions (i.e., the Neale; see Madelaine & Wheldall, 1998) fell somewhere between the two researcher-made questions (.71). Some problems with questions may explain the variability in the correlations. These include the difficulty of the questions, or the extent to which answers may have been apparent without reading the passage. Although this cannot be determined without examining the passages and questions, either phenomenon may explain the low correlations reported by Markell.

The third highest mean correlation (.69) was found when standardized, group-administered tests were used as the criterion. While many different types of group-administered tests were used, they all required the student to perform a similar task: read passages and answer multiple-choice questions based on what had been read. Although the group-administered tasks were similar, the correlations for each test ranged from .51 to .85. Most notably, among these standardized, group-administered tests, the California Achievement Test resulted in the lowest figures, with an average correlation of .57. By eliminating the four correlations reported in one study using the California Achievement Test, the mean correlation for group-administered criterion tests increased from .69 to .74.

Although many of the criterion measures were individually administered, the Kaufman Test of Educational Achievement-Brief did not fit into the other categories already discussed. The Kaufman Test of Educational Achievement- Brief is unique because of the tasks required of the student. First, the student identifies letters and words, next they act out what is read. The Kaufman Test of Educational Achievement- Brief provided one of the lowest correlations (.41), which was significantly different from mazes (.78), Z

(1, 267) = 3.99, $p < .001$, basal reading tests (.71), $Z(1, 195) = 2.81$, $p < .01$, questions on passages (.71), $Z(1, 105) = 2.26$, $p < .05$, and standardized, group-administered tests (.69), $Z(1, 1501) = 2.97$, $p < .01$. While the Kaufman Test of Educational Achievement-Brief criterion task was unique, it is difficult to determine what explains the low correlation because no other studies used the Kaufman Test of Educational Achievement- Brief or similar tasks.

Text fluency and grade. Another variable worth looking at is the students' grade (see Table 6). In the included studies, students' grade ranged from 1 through 8. The majority of studies (75%) provided correlations based on students' individual grade level; the remaining studies (25%) combined three or more grade levels.

Studies by Jenkins and Jewell (1993) and Kranzler, Miller, and Jordan (1999), which looked at correlations between text fluency and different criterion measures across grades, found that correlations were higher in lower grades (i.e., second, third, and fourth grade), and then dropped off after fifth grade. This same trend is evident in Table 6. One explanation is that students in the higher grades read more expository material, requiring them to focus more on the ideas and content of the material (Shinn & Hubbard, 1992). For example, textbooks in the higher grades contain facts that students may need to remember whereas textbooks in the lower grades, which primarily tell stories, are potentially more engaging. Another explanation is that words read correct represents basic skills that more clearly relate to comprehension, as indexed in the lower grades.

Therefore, based on these explanations, it is not surprising that mean correlations for second through fourth grades were for the most part larger than fifth and sixth grade. The exception to this trend was at first grade, which had a low correlation similar to that of sixth grade. However, the correlation for first grade (.62), was only significantly different from the mean correlation reported for second grade

(.75), $Z(1, 451) = 1.82, p < .05$. The low correlation reported for first grade is not unexpected given that first graders are typically just beginning to read; thus, their skills cause a floor effect. Similar to the ceiling effect, described earlier, the floor effect occurs when the range of difficult items is restricted so that many of the students earn a low score. To exacerbate this problem, this first-grade sample was from a Title 1 school. It may be that fluency is not highly correlated with comprehension for at-risk children just learning to read. Perhaps prereading skills, such as identification of letters and letter-sound correspondences, are more appropriate to measure for students at this level. Moreover, it may not be appropriate to measure reading comprehension for first-grade students given their limited word-reading capacity.

Although the average correlations were highest in second through fourth grade, the strongest correlation was in second grade (.75), which was significantly different compared to third grade (.68), $Z(1, 908) = 2.14, p < .05$, fourth grade (.68), $Z(1, 875) = 2.11, p < .05$, fifth grade (.59), $Z(1, 696) = 3.86, p < .001$, and sixth grade (.62), $Z(1, 578) = 2.80, p < .01$. Third and fourth grade had the same mean correlation (.68), which was the second highest correlation. Third grade was significantly higher than fifth grade (.59), $Z(1, 832) = 2.10, p < .05$; and fourth grade was significantly higher than fifth grade (.59), $Z(1, 799) = 2.07, p < .05$.

The remaining six studies combined students across grades, with the highest correlation reported for studies that grouped grades 1-6 (.81), and 4-8 (.81). Although the correlations are the same, grouping grades one through six provided a significantly higher correlation compared to the following individual grade levels; first grade (.62), $Z(1, 301) = 2.83, p < .01$; second grade (.75), $Z(1, 622) = 1.86, p < .05$; third grade (.68), $Z(1, 758) = 3.79, p < .001$; fourth grade (.68), $Z(1, 725) = 3.75, p < .001$; fifth grade (.59), $Z(1, 546) = 5.18, p < .001$; and sixth grade (.62), $Z(1, 428) = 4.12, p < .001$. While grouping grades 4-8 was only significantly higher

than first grade (.62), $Z(1, 99) = 1.85, p < .05$; fifth grade (.59), $Z(1, 344) = 2.42, p < .01$; and sixth grade (.62), $Z(1, 226) = 2.10, p < .05$, it was not significantly different from second grade (.75), $Z(1, 417) = .84, ns$; third grade (.68), $Z(1, 556) = 1.64, ns$; and fourth grade (.68), $Z(1, 517) = 1.63, ns$.

While critics claim that combining grades creates an artificially heterogeneous group and inflates estimates (Mehrens & Clarizio, 1993), some additional points are worth noting. The highest correlations reported (.91), not only for these five studies but also for all 33 correlations, was by Fuchs (1981; see also Fuchs & Deno, 1992) and Fuchs et al. (1988). These two studies are unique for various reasons. The first study by Fuchs (1981) looked at two different types of basal reading series to determine if passage source (i.e., the students' basal reader or a different basal reading series) or passage difficulty affected the correlation between text fluency and comprehension. The correlations were just as strong for the different sources as they were for difficulty level (.91).

The second study by Fuchs et al. (1988) is unique because it is one of only two studies where the sample was comprised entirely of students with disabilities. In addition, four separate criterion measures were used: the Stanford Achievement Test (.91), questions on the passages (.84), cloze technique (.75), and story retell (.74). Interestingly, the correlations for the Stanford Achievement Test did not follow the trends identified under criterion measures already discussed. This indicates that the student group (i.e., general, special, combination of general and special education, and at-risk) may impact the magnitude of the correlations reported.

Text fluency and student group. Of these included studies, 58% used students in general education; 33% used a combination of students with and without disabilities; 6% used students with a disability; and 12% used students identified as at-risk (e.g., Title 1 or

lowest readers). Table 7 provides correlations by student group. Correlations based on special education students were significantly higher (.77) from those based on general education students (.66), $Z(1, 1338) = 2.48, p < .01$, and from those based on students at-risk (.66), $Z(1, 264) = 1.83, p < .05$. In addition, studies that combined students in general and special education placements reported a significantly higher mean correlation (.71) than studies only using students in general education (.66), $Z(1, 2081) = 2.12, p < .05$. However, only two studies used students with disabilities (Lomax, 1983; Fuchs et al., 1988), and only two studies used students at-risk (Bain & Garlock, 1992; Parker, Hasbrouck, & Tindal, 1992) making conclusions tenuous. Therefore, it is important to look at all of the variables in order to determine which factors impact the magnitude of the correlation. The last variable to be examined under text fluency is sample size.

Text fluency and sample size. Of the correlations, 39% were run with an n under 50; 39% used an n between 50 and 99; and the remaining 22% had an n greater than 99 (see Table 8). A small sample size may affect the reliability of the correlation, meaning the chances of obtaining a similar score with the same population decreases as sample size decreases. Therefore, it is desirable to obtain the largest sample size feasible. For correlational research, a minimum of 30 subjects is typically considered necessary (Gall, Borg, & Gall, 1996).

Although the cutoffs described above are arbitrary, they allow for a comparison of small, medium, and larger sample size studies. As indicated in Table 8, those studies with a sample size below 50 reported the largest mean correlation (.73). There was a significant difference between the mean correlation reported for studies with a sample size below 50 and those with a sample size between 50 and 99 (.65), $Z(1, 1357) = 2.66, p < .01$. However, the mean correlation reported for studies with sample sizes below 50

was not significantly different than studies using sample sizes greater than 99 (.70), $Z(1, 1434) = 1.08$, *ns*. In addition, the mean correlations for studies using sample sizes greater than 99 (.70) were significantly different from studies using sample sizes between 50 and 99 (.65), $Z(1, 1901) = 2.00$, $p < .05$. This indicates that studies with a sample size below 50 and greater than 99 both provided higher correlations than studies using a sample size between 50 and 99. However, the mean correlation provided by studies using a sample size below 50 is less reliable.

Correlations for List Fluency

The word lists were derived from three different sources. The majority (71%) used basal readers (i.e., Ginn; Holt; Houghton-Mifflin; Harris-Jacobson; or Scott-Foresman), with a mean correlation of .61 ($SD = .07$). One study (14%) used a word list from the Basic Elementary Reading Vocabulary-R series, with a correlation of .73; and the other study (14%) used words from a passage, randomly placing them to form a word list, with a correlation of .53. The studies using the Vocabulary-R series provided a significantly different correlation (.73) than the one study that used words from a passage (.53), $Z(1, 156) = 1.87$, $p < .05$. Although this finding may be of interest, it does not fully explain the relation between list fluency and comprehension. To examine this relation further, correlations were examined by number of words on list, list level, criterion measure, grade, student group, and sample size.

List fluency and length. The word lists ranged in length from 60 to 400 words. The length of the word list may affect the correlation between list fluency and comprehension, because of the need to prorated scores on shorter lists. As already indicated, prorating may affect the correlation by increasing the spread and increasing the correlations. Although not one study reported the use of prorating, it is possible this occurred with lists as short as 60 or 150 words.

Table 9 provides a breakout of the studies grouped by length of word list. The highest correlation (.73) occurred for studies using the shortest list (60 words) while the lowest correlation (.53) was reported for studies using the longest list (400 words). The mean correlation provided for the 60 word list was significantly different from the 400-word list, $Z(1, 156) = 1.87$ $p < .05$. Although prorating may explain differences in the size of the correlation, no definitive conclusions can be drawn because the authors did not report whether prorating was used. Therefore, other factors, such as the level of the words on the lists (i.e., below, at, or above grade level), should be considered.

List fluency and difficulty level. Another factor that may affect the correlation between list fluency and comprehension is whether the difficulty of the list is below, at, or above the students' grade level. As already mentioned, students reading below grade level could experience a ceiling effect; this would result in a nonnormal distribution. To determine if the difficulty of the word lists was a factor, correlations for the studies using word lists below, at, and above grade level were compared. Only 5 of the 7 correlations could be used because two studies did not provide this information (see Table 10). Moreover, only one correlation was reported for a word list below grade level (.53). By contrast, four correlations were provided for word lists at grade level (.62), and no correlation was reported for above grade level lists. The difference between the correlation on a word list below grade level (.53) was not significantly different from those at grade level (.62), $Z(1, 251) = 1.05$, ns . Of course, with only one correlation for below grade level word lists, it is not defensible to draw any conclusions.

List fluency and criterion measure. Although four different criterion measures were used in these studies, two of the four had dependency in their data and were, therefore, eliminated from this analysis (see Table 11). Of the two remaining measures, basal reading

series (the Ginn; Holt; Houghton-Mifflin; or Scott-Foresman) were used the most (67%). The remaining measure (33%) was a standardized, group-administered test (e.g., Iowa Test of Basic Skills). There was no significant difference between the basal reading series (.62) and the standardized, group-administered test (.55), $Z(1, 334) = .957$, ns. There was a significant difference, however, between the mean correlation for standardized, group-administered tests for text fluency (.69) and list fluency (.55), $Z(1, 1640) = 3.00$, $p < .001$. However, there was no difference for basal reading series between text fluency (.71) and list fluency (.62), $Z(1, 278) = 1.34$, ns.

List fluency and grade. Subjects in these studies ranged from first through sixth grade. Seventy-one percent of the correlations were provided for students in grades 4 through 6. The remaining correlations (29%) combined four or more grade levels. As shown in Table 12, not all grades are represented. In addition, fifth and sixth grade each had only one correlation, making it difficult to compare grades. There was no significant difference between the correlations at fourth, fifth and sixth grade. The difference between fourth grade (.56) and fifth grade (.69) was, $Z(1, 204) = .04$, ns, while fourth (.56) and sixth grade was (.55), $Z(1, 204) = .09$, ns. In addition, the difference between fifth grade (.69) and sixth grade (.55) was, $Z(1, 92) = 1.08$, ns.

Moreover, there was no significant difference between studies that combined grade levels 1-6 (.73) and those that combined grades 3-6 (.58), $Z(1, 126) = 1.39$, ns. However, there was a significant difference between the combined grades 1-6 (.73) and fourth grade (.56); $Z(1, 202) = 1.70$, $p < .05$. In comparison to text fluency, where combining grades 1-6 provided a significant difference with each grade (one through six), combining grades 1-6 on list fluency provided only one significant difference with fourth grade. This is not surprising, considering list fluency provided only three separate grades (4 through 6) whereas text fluency had six separate grades (1 through 6).

List fluency and student group. Of these included studies, 57% used students in general education; 29% used a combination of students in general and special education; and one study (14%) used students at-risk. Table 13 provides an outline of these studies based on student group. Unlike findings from the studies comparing text fluency, there was no difference between these groups based on type of student group. The mean correlation for students in general education (.62), was not significantly different from students in general and special education (.59), $Z(1, 296) = .40$, ns, or between students at-risk (.58), $Z(1, 221) = .45$, ns. Similarly, there was not a significant difference between the mean correlation for students in general and special education (.59) with students at-risk (.58), $Z(1, 239) = .12$, ns. This indicates that student group did not affect the strength of the correlation. In addition to the variables discussed, sample size was also looked at to determine what, if any, impact this had on the reported correlations.

List fluency and sample size. To determine if sample size was a factor, correlations were broken into those having a sample size below 50 (71%), those with an n between 50 and 99 (14%), and those with an n greater than 99 (14%). As already mentioned, these cutoffs are arbitrary, but they do allow for a comparison of small, medium, and larger sample size correlations.

Table 14 shows correlations by sample size. Similar to those findings reported under criterion measures and grade, there was no difference between the mean correlations reported for sample size. The difference between a sample size below 50 (.65) was not significantly different from a sample size between 50 and 99 (.58), $Z(1, 266) = .84$, ns, and a sample size greater than 99 (.53), $Z(1, 296) = 1.53$, ns. Additionally, the difference between sample sizes between 50 and 99 (.58) was not significantly different from sample sizes greater than 99 (.53), $Z(1, 194) = .49$, ns. However, with the small number of correlations reported for list fluency, it is difficult to determine what variables, if any, affected the correlations. The information from one study may only generalize to populations that are

similar and that use the same criterion measures and word lists. Therefore, it is important to consider the limitations of the studies while summarizing the information based on the mean correlations.

Correlations for Text Accuracy

The passages used in these studies were derived from two sources, basal readers or written by the authors. The majority of the correlations were provided using passages from basal readers (80%), with a mean correlation of .71 ($SD = .19$). The remaining correlation used passages written by the author (20%), with a correlation of .50. The difference between the correlations from these two types of passages was significant $Z(1, 266) = 2.70, p < .01$. However, with only one correlation provided for passages written by the author, it is not possible to make generalizations. Although the small number of correlations is problematic for drawing conclusions an analysis was conducted in terms of words per passage, criterion measure, grade and sample size. This provided comparisons between the five correlations reported as well as allowing for comparisons between correlations for the other word reading measures (i.e., text and list fluency).

Text accuracy and length. The length of the passages ranged from 100 to 400 words. Twenty percent of the correlations used 100-word passages; 60% used 250-word passages; and another 20% used 400-word passages. Table 15 provides a breakdown of the correlations by passage length. The shortest passages (100 words) had the highest mean correlation (.84), which was significantly different from the 250-word passages (.53), $Z(1, 153) = 3.79, p < .001$, and the 400-word passages (.50), $Z(1, 202) = 4.70, p < .001$. A similar trend was evident in the mean correlations for text fluency, where the 100-word passages provided the highest mean correlation (.91) compared to all other passage lengths. While prorating was considered a potential problem for passages and word

lists used for text and list fluency, this problem is not reflected in the percentage of words read correctly, because prorating would not be used. However, shorter passage lengths may be problematic because students are reading fewer words, which may decrease their chances of making errors.

Text accuracy and criterion measures. Two criterion measures were used: a standardized group-administered test (Iowa Test of Basic Skills), and maze passages (see Table 16). A total of four correlations (80%) were provided for the maze, with a mean correlation of .71 ($SD = .19$). Only one correlation (20%), was provided for a standardized, group-administered test (.50). The difference between the correlations on these two criterion measures was significant $Z(1, 266) = 2.70, p < .01$. This is similar to the findings for text fluency where studies using mazes provided the highest correlations. Unlike text fluency, the small number of correlations provided for text accuracy does not leave room for interpretation. However, analyses were conducted on the correlations based on grade, student group, and sample size which not only allowed for comparisons in each of these areas, but across the different types of word-reading measures as well.

Text accuracy and grade. Students' grade ranged from 1 through 6. The majority of studies (80%) provided correlations based on students' individual grade level; the remaining study (20%) combined grades 1-6. As indicated in Table 17, sixth grade provided a correlation of .71, which was significantly different from the correlation at fifth grade (.30), $Z(1, 40) = 1.73, p < .05$. The one study that combined grades 1-6 reported the highest correlation. There was a significant difference between the combined grades correlation of .84 and fourth grade (.51), $Z(1, 224) = 4.78, p < .001$, fifth grade (.30), $Z(1, 110) = 3.52, p < .001$, and sixth grade (.71), $Z(1, 110) = 2.22, p < .01$. As already mentioned, combining grades can impact the correlation by creating an artificially heterogeneous group. It

is interesting to note that the highest correlation, based on grade, for text fluency, list fluency, and for text accuracy have all been from combining grades 1-6.

Text accuracy and student group. Of these included studies, 40% used a combination of students in general and special education; and 60% used students at-risk (see Table 18). There were no studies that reported using only students in general education or only students in special education. There was no significant difference between the combined group of general and special education students (.65) and the group of students at-risk (.53), $Z(1, 266) = 1.27, ns$. These results are similar to those studies reported under list fluency in that there was no significant difference between these groups based on student group. In contrast, students in special education provided the highest correlation for text fluency.

Sample size. The number of subjects in the studies ranged from 21 to 113. Sixty percent of the correlations were derived with an n under 50; one correlation (20%), with an n between 50 and 99; and one correlation (20%), with an n greater than 99. The largest correlation (.84) was for the study using between 50 and 99 subjects (see Table 19). The correlation for the study using between 50 and 99 subjects was significantly higher than the correlation for studies using below 50 subjects (.53), $Z(1, 153) = 3.79, p < .001$, and the one using greater than 99 subjects (.50), $Z(1, 202) = 4.70, p < .001$.

These mean correlations differ from those reported for text fluency. There, the mean correlations for sample sizes below 50 (.73) and greater than 99 (.70) were both significantly different than those studies using a sample size between 50 and 99 (.65). In comparison, the correlations reported for list fluency indicated no significant differences between the three different sample sizes. As

indicated previously, the small number of correlations used should be taken into consideration when making generalizations about these results.

Correlations for List Accuracy

The word lists were derived from two different sources. The majority of correlations (86%) used lists from commercially developed tests (i.e., Diagnostic Reading Scales; Gilmore Oral Reading Test; Science Research Associates Achievement Series; Slosson Oral Reading Test; Woodcock Reading Mastery Test; Wide Range Achievement Test- Revised; Wechsler Objective Reading Dimensions; or Woodcock-Johnson Psycho-educational Battery- Revised), with a mean correlation of .81 ($SD = .10$). The remaining comparison (14%) used an author- made measure (i.e., Comprehensive Reading Assessment Battery) to form the list, with a correlation of .61. There was a significant difference between the commercially developed tests and the one correlation using the Comprehensive Reading Assessment Battery, $Z(1, 935) = 4.12, p < .001$. Similar to the areas discussed previously, the mean correlation of these measures was broken down further (using number of words per list, criterion measure, grade, student group, and sample size) to examine the impact these variables may have had on the magnitude of the correlation. In addition, these breakdowns allow comparisons with the other measures of word reading.

List accuracy and length. Word list length was reported for only two of the seven correlations (see Table 20). A study conducted by (Rice, 1981), which looked at the strength of the relation between word recognition on word lists and reading comprehension for first graders, used a list of 20 words. The second study by Jenkins et al. (2000), which looked at how accuracy and fluency in passages and on word lists relate to reading comprehension, used a list of 400 words. There was no significant difference

between these two correlations based on the word list length, $Z(1, 217) = .88$, ns. This is in contrast to list fluency, which did produce a significant difference between the 60-word lists (.73) and the 400-word lists (.53).

List accuracy and criterion measure. Five criterion measures were used, although two of the five had dependency in their data so only three are included in this analysis. The most common measure was standardized, group-administered tests (67%) (i.e., Canadian Test of Basic Skills, Cooperative Primary Test, Comprehensive Tests of Basic Skills, Diagnostic Reading Scales, Iowa Test of Basic Skills). One correlation each (17%) used the Woodcock-Johnson Psycho-Educational Battery- Revised, and story retell with additional comprehension questions asked following the story retell.

The highest mean correlation was associated with the Woodcock-Johnson Psycho-Educational Battery- Revised (see Table 21). The correlation provided for this criterion measure (.89) was significantly different from the standardized, group-administered test (.67), $Z(1, 725) = 8.20$, $p < .001$, and the story retell with comprehension questions (.60), $Z(1, 385) = 3.39$, $p < .001$. However, there was no difference between the standardized, group-administered tests (.67) and story retell with comprehension questions (.60), $Z(1, 390) = .55$, ns. In addition to looking at the impact the criterion measures had on the correlations, grade, student group and sample size were examined as well.

List accuracy and grade. Students' grade ranged from 1 through 6 (see Table 22). Four correlations (57%) were based on students at one grade level while the remaining three correlations (43%) combined students from two or more grades. Although there was no significant difference between grades 1, 2, and 4, there was a significant difference between studies that grouped grades with the correlations provided for first, second, and fourth grades. The difference was significant for the study that grouped grades 1-6 (.81)

with first grade (.68), $Z(1, 205) = 2.11, p < .05$, second grade (.54), $Z(1, 171) = 3.33, p < .001$, and fourth grade (.61), $Z(1, 212) = 3.01, p < .01$. The difference was also significant for the study that grouped grades 2-3 (.82) with first grade (.68), $Z(1, 288) = 2.66, p < .01$, second grade (.54), $Z(1, 254) = 3.91, p < .001$, and fourth grade (.61), $Z(1, 295) = 3.71, p < .001$. For the study that grouped grades 2-4 (.89) the difference was again significant for first grade (.68), $Z(1, 465) = 5.30, p < .001$, second grade (.54), $Z(1, 431) = 6.22, p < .001$, and fourth grade (.61), $Z(1, 472) = 6.54, p < .001$.

In comparison with the other three types of word-reading measures, it is apparent that combining grades did have an impact on the correlations. The highest correlation provided by all four types of word-reading measures were with studies that combined grades. In contrast, a comparison of all four word-reading measures based on a single grade level revealed only two significant differences. At second grade, text fluency provided a higher correlation (.75) than list accuracy (.54), $Z(1, 457) = 2.82, p < .01$; and at fourth grade, text fluency provided a higher correlation (.68) than text accuracy (.51), $Z(1, 623) = 1.76, p < .05$. Because some areas did not provide correlations for grade levels one through six, it is not possible to determine clear differences between all of these grades and the four types of word-reading measures.

List accuracy and student group. Of these included studies, 43% used students in general education; 14% used students in special education; 43% used a combination of students in general and special education, while no studies reported using students at-risk (see Table 23). There was no significant difference between the group of students in general education (.75) with the group of students in special education (.81), $Z(1, 415) = 1.33, ns$, and the group of students in general and special education (.79), $Z(1, 834) = .73, ns$. Moreover, there was no significant difference between the group of students in special education (.81) and the group of

students in general and special education (.79), $Z(1, 619) = .51$, *ns*. These results are similar to the comparisons made using list fluency and text accuracy, where there was no significant difference based on student group. In contrast, text fluency did provide a difference where students in special education provided the highest correlation.

List accuracy and sample size. The number of subjects in the studies ranged from 26 to 61. Twenty-nine percent of the correlations were derived with an n under 50; no correlations reported using an n between 50 and 99; and five correlations (71%) were based on n 's greater than 99 (see Table 24). There was a significant difference between correlations using an n below 50 (.54) and those using an n greater than 99 (.80), $Z(1, 935) = 3.96$, $p < .001$. However, it is difficult to determine what impact sample size had on the correlations due to the small number of correlations provided and because there were no studies that used a sample size between 50 and 99.

Conclusions

Results of this review indicate that the relation between word-reading skills and an individual's capacity to derive meaning from text is strong, regardless of how word-reading skill is indexed. For each of the four word-reading measures (i.e., text fluency, list fluency, text accuracy, and list accuracy), correlations with measures of reading comprehension were high. Unfortunately, due to dependency in the data across word-reading measures, it was not possible to test the overall correlations among the four word-reading measures. In addition, definitive conclusions cannot be drawn related to the impact specific variables had on the correlations across the four types of word-reading measures, due to dependency in the data and because some types of word-reading measures did not

provide correlations in some areas (e.g., not all grades were represented across all four types of word-reading skill). Nevertheless, some points are worth noting related to the variables that were observed for each type of word-reading measure.

The first variable examined across all four types of word reading was the length of passages or word lists. This is important especially for text fluency and list fluency because the use of prorating may affect correlations by inflating the spread of scores. Interestingly, shorter text passages (i.e., 100 words) and shorter lists (i.e., 60 words) produced higher correlations for text fluency and list fluency, respectively. In addition, shorter passages for text accuracy (100 words) also produced higher correlations. While prorating would not affect an accuracy score, fewer words may not provide enough information for determining students' word-reading skills. One way to determine the effects of prorating on the size of the correlation between word-reading measures and reading comprehension is to conduct a study specifically designed for this purpose, which carefully controls all other design features. Unfortunately, no such study was identified.

The second variable examined across all four types of word reading was the type of criterion measure used. However, it was difficult to compare criterion measures across all the word-reading measures because a limited number of criterion measures (2 to 3) were included for three of the measures (i.e., list fluency, text accuracy, list accuracy), while five were used for text fluency. Nevertheless, the highest correlation was reported for mazes used with text fluency and text accuracy. As discussed previously, this is not surprising given the similarities between these two tasks (i.e., they both require students to read passages). It would be interesting to know if the maze criterion would have also produced the highest correlations for list fluency and list accuracy. However, it seems unlikely because the task of reading a word list and a maze passage are not as similar as reading a text passage and a maze passage.

Student grade level was the third area that was examined across all four measures of word reading. While Jenkins and Jewell (1993) and Kranzler et al. (1999) found that the correlations across grades between text fluency and a measure of reading comprehension were typically higher for younger grades, similar trends were not observable for the other three types of word-reading measures because not all of the grades were represented. For example, list fluency and text accuracy only provided information for grades 4 through 6; list accuracy only had information on grades 1, 2, and 4.

However, it was possible to compare grades that were represented in two or more of the four word-reading measures to determine if a particular area yielded higher correlations. The only difference observed was for text fluency, which provided a higher correlation than list accuracy for second grade as well as a higher correlation than text accuracy for fourth grade. Otherwise, there was no difference between the types of word-reading measures based on grade level. Although text fluency provided a higher correlation at second grade compared to list accuracy, no studies were found that used list fluency or text accuracy; therefore, no clear conclusions can be drawn. However, all four types of word reading were used at fourth grade with text fluency providing a higher correlation than text accuracy. Nevertheless, list fluency, list accuracy, and text fluency all provided similar correlations. These results also make it difficult to determine clear distinctions across word-reading measures based on grade. Unfortunately, no studies were found that looked at the different types of word-reading measures across grades.

The fourth area that was examined across all four types of word-reading measures was student group. As was the case with the exclusion of grades in some types of word reading, the exclusion of various student groups made it difficult to determine what impact this variable had on the correlations. Student group did not make a difference for three of the four types of word reading. The only

difference observed was for text fluency where students in special education had a significantly higher mean correlation compared to students in general education or students at risk. Unfortunately, students in special education were not included for studies looking at list fluency or text accuracy, therefore making it difficult to compare student group across all four types of word-reading measures. One way to determine if student group impacts the magnitude of the correlation is to compare different types of student groups on the same measures of word reading. However, no studies specifically addressed this; and without this information, it is not possible to draw conclusions.

The last variable that was examined across the four word-reading measures was sample size. While there was no difference in the correlation for list fluency as a function of sample size, there were differences for text fluency, text accuracy, and list accuracy. Moreover, the three types of word-reading measures for which correlations were reported as a function of sample size all indicated a different sample size as the highest correlation. For example, text fluency indicated below 50 and above 99 were both higher than studies using a sample size between 50 and 99; whereas text accuracy indicated a sample size between 50 and 99 was the highest; and list accuracy indicated a sample size above 99 was the highest. Therefore, it appears that sample size did not have a consistent effect on all of the word-reading measure indices. Unfortunately, similar to the other variables, it is not possible to draw conclusions due to the lack of consistency across the four types of word-reading measures.

In conclusion, due to the number of students who struggle with reading, and the focus teachers and researchers place on measuring word-reading skill, it seems important to determine what variables if any impact the relation between measures of word-reading and comprehension. While this review revealed trends for some variables, the most sensible conclusion at this time is that

more research needs to be conducted. To help answer the question about the effects of prorating, future research should focus on the length of passages (e.g., 100 words vs. 400 words) and word lists (e.g., 60 words vs. 400 words) and the effect length has on the correlation with comprehension. In addition, although Fuchs et al. (1988) reported correlations between word-reading with four different measures of reading comprehension on students in special education, this research should be extended to other populations to determine how criterion measures affect the correlation with populations other than students in special education. Although grade seemed to impact the correlations for text fluency with correlations being higher in younger grades, future research should extend this using list fluency, along with text and list accuracy. This would help determine if different measures of word-reading skill have higher correlations with comprehension for different grades. Last, the impact student group has on the magnitude of the correlation could be addressed by conducting research using the same measures of word reading and comprehension with different groups of students to determine if there are differences.

References

- Aaron, P. G., Joshi, M., & Williams, K. A. (1999). Not all reading disabilities are alike. Journal of Learning Disabilities, *32*, 120-137.
- Adams, M. J. (1990). Beginning to read. Cambridge MA: MIT Press.
- Armstrong, S. W. (1983). The effects of material difficulty upon learning disabled children's oral reading and reading comprehension. Learning Disability Quarterly, *6*, 339-348.
- Bain, S. K., & Garlock, J. W. (1992). Cross-validation of criterion-related validity for CBM reading passages. Diagnostic, *17*, 202-208.
- Biemiller, A. (1977-1978). Relationship between oral reading rates for letters, words, and simple text in the development of reading achievement. Reading Research Quarterly, *2*, 223-253.
- Bowers, P. G., & Swanson, L. B. (1991). Naming speed deficits in reading disability: Multiple measures of a singular process. Journal of Experimental Child Psychology, *51*, 195-219.
- Buehner, L. J. (1983). The perception of prosodic cues and its relationship to reading fluency. (Doctoral dissertation, The University of Rochester, 1983). Dissertation Abstracts International, *44*, ATT 8321706.
- Collins, V. L. (1989). Reading fluency: An examination of its role in the reading process. Unpublished master's thesis, University of Oregon, Eugene.
- Curtis, M. E. (1980). Development of components of reading skill. Journal of Educational Psychology, *72*, 656-669.

Deno, S. L., Mirkin, P. K., & Chiang, B. (1982). Identifying valid measures of reading. Exceptional Children, 49, 36-45.

DeSoto, J. L., & DeSoto, C. B. (1983). Relationship of reading achievement to verbal processing abilities. Journal of Educational Psychology, 75, 116-127.

Ehri, L., & Saltmarsh, J. (1995). Beginning readers outperform older disabled readers in learning to read words by sight. Reading and Writing: An Interdisciplinary Journal, 7, 295-326.

Fuchs, L. S. (1981). The concurrent validity of progress measures of basal reading material. (Doctoral dissertation, University of Minnesota, 1981). Dissertation Abstracts International, 42, AAT 8206352.

Fuchs, L. S., & Deno, S. L. (1992). Effects of curriculum within curriculum-based measurement. Exceptional Children, 58, 232-243.

Fuchs, L. S., Fuchs, D., Hosp. M. K., & Jenkins, J. (2000). Oral reading fluency as an indicator of reading competence: A theoretical, empirical, and historical analysis. Manuscript submitted for publication.

Fuchs, L. S., Fuchs, D., & Maxwell, L. (1988). The validity of informal reading comprehension measures. Remedial and Special Education, 9, 20-28.

Fuchs, L. S., Tindal, G., & Deno, S. L. (1984). Methodological issues in curriculum-based reading assessment. Diagnostic, 9, 191-207.

Fuchs, L. S., Tindal, G., Fuchs, D., Shinn, M., Deno, S., & Germann, G. (1983). The technical adequacy of a basal reading mastery test: The Holt basic reading series (Research Report No. 130). Minneapolis: University of Minnesota Institute for Research on Learning Disabilities.

Fuchs, L. S., Tindal, G., Shinn, M., Fuchs, D., Deno, S., & Germann, G. (1983). Technical adequacy of a basal readers' mastery tests: The Ginn 720 series (Research Report No. 122). Minneapolis: University of Minnesota Institute for Research on Learning Disabilities.

Gall, M. D., Borg, W. R., & Gall, J. P. (1996). Educational research an introduction. White Plains, NY: Longman.

Gough, P. B., Hoover, W., & Peterson, C. L. (1996). Some observations on the simple view of reading. In C. Cornoldi & J. Oakhill (Eds.), Reading comprehension difficulties (pp. 1-13). Mahwah, NJ: Erlbaum.

Hintze, J. M., Shapiro, E. S., Conte, K. L., & Basile, I. M. (1997). Oral reading fluency and authentic reading material: Criterion validity of the technical features of CBM survey-level assessment. School Psychology Review, 26, 535-553.

Hoover, W. A., & Gough, . B. (1990). The simple view of reading. Reading and Writing: An Interdisciplinary Journal, 2, 127-160.

Jenkins, J. R., Fuchs, L. S., Espin, C., Broek, P., & Deno, S. L. (2000). Effects of task format and performance dimension on word reading measurers: Criterion validity, sensitivity to impairment, and context facilitation. Manuscript submitted for publication.

Jenkins, J. R., & Jewell, M. (1993). Examining the validity of two measures for formative teaching: Reading aloud and maze. Exceptional Children, 59, 421-432.

Kranzler, J. H., Brownell, M. T., & Miller, M. D. (1998). The construct validity of curriculum-based measurement of reading: An empirical test of a plausible rival hypothesis. Journal of School Psychology, 36, 399-415.

Kranzler, J. H., Miller, M. D., & Jordan, L. (1999). An examination of racial/ethnic and gender bias on curriculum-based measurement of reading. School Psychology Quarterly, 14, 327-342.

LaBerge, D., & Samuels, S. J. (1974). Toward a theory of automatic information processing in reading. Cognitive Psychology, 6, 293-323.

Lesgold, A. M., & Perfetti, C. A. (1978). Interactive processes in reading comprehension. Discourse Processes, 1, 323-336.

Lomax, R. G. (1983). Applying structural modeling to some component processes of reading comprehension development. Journal of Experimental Education, 52, 33-40.

Lopez, L. C., Clark, P. M., & Winer, G. A. (1979). The relationship of two cognitive styles to reading speed and comprehension. The Journal of Genetic Psychology, 135, 229-235.

Lyon, G. R. (1997, July). Statement on reading. Report presented before the committee on Education and the Workforce U.S. House of Representatives, Washington, D.C.

Madelaine, A., & Wheldall, K. (1998). Towards a curriculum-based passage reading test for monitoring the performance of low-progress readers using standardized passages: A validity study. Educational Psychology, 18, 471-478.

Markell, M. A. (1991). Contributions of oral reading fluency to measures of reading comprehension. (Doctoral dissertation, University of Minnesota, 1991). Dissertation Abstracts International, 52, AAT 9209433.

Marston, D. B. (1982). The technical adequacy of direct, repeated measures of academic skills in low achieving elementary students. (Doctoral dissertation, University of Minnesota, 1982). Dissertation Abstracts International, 43, AAT 8301966.

Marston, D. B., & Deno, S. (1982). Implementation of direct and repeated measurement in the school setting (Research Report No. 106). Minneapolis: University of Minnesota Institute for Research on Learning Disabilities.

McCormick C., & Samuels, S. J. (1979). Word recognition by second graders: The unit of perception and interrelationships among accuracy, latency, and comprehension. Journal of Reading Behavior, 11, 107-118.

Mehrens, W. A., & Clarizio, H. F. (1993). Curriculum-based measurement: Conceptual and psychometric considerations. Psychology in the Schools, 30, 241-254.

Nathan, R. G., & Stanovich, K. E. (1991). The causes and consequences of differences in reading fluency. Theory Into Practice, 30, 176-184.

Nation K., & Snowling, M. (1997). Assessing reading difficulties: The validity and utility of current measures of reading skill. British Journal of Educational Psychology, 67, 359-370.

Parker, R., Hasbrouck, J. E., & Tindal, G. (1992). Greater validity for oral reading fluency: Can miscues help? The Journal of Special Education, 25, 492-503.

Perfetti, C., Finger, E., & Hogaboam, T. (1978). Sources of vocalization latency differences between skilled and less skilled young readers. Journal of Educational Psychology, 70 (5), 730-739.

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

Perfetti, C. A., & Hogaboam, T. (1975). Relationship between single word decoding and reading comprehension skill. Journal of Educational Psychology, 67, 461-469.

Posner, M. I., & Snyder, C. R. R. (1975a). Attention and cognitive control. In R. Solso (Ed.), Information processing and cognition: The Loyola symposium. (pp. 55-85). Hillsdale, NJ: Erlbaum.

Posner, M. I. & Snyder, C. R. R. (1975b). Facilitation and inhibition in the processing of signals. In P. M. A. Rabbitt & S. Dornic (Eds.), Attention and performance (vol.5, pp. 669-682). New York: Academic Press.

Rasinski, T. V. (1990). Investigating measures of reading fluency. Educational Research Quarterly, 14 (3), 35-44.

Rice, N. K. (1981). The relationship between oral reading fluency and other reading behaviors among first grade children. (Doctoral dissertation, Michigan State University, 1981). Dissertation Abstracts International, 42, ATT 8253615.

Roberts, M., & Smith, D. D. (1980). The relationship among correct and error oral reading rates and comprehension. Learning Disabilities Quarterly, 3, 54-64.

Samuels, S. J. (1987). Information processing abilities and reading. Journal of Learning Disabilities, 20, 18-22.

Samuels, S. J., & Flor, R. F. (1997). The importance of automaticity for developing expertise in reading. Reading & Writing Quarterly: Overcoming Learning Difficulties, 13, 107-121.

Shankweiler, D., & Liberman, I. Y. (1972). Misreading: A search for causes. In J. Kavanagh & I. Mattingly (Eds.), Language by ear and eye. Cambridge, MA: MIT Press.

Shankweiler, D., Lundquist E., & Katz, L. (1999). Comprehension and decoding: Patterns of association in children with reading difficulties. Scientific Studies of Reading, 3, 69-94.

Shinn, M. R., & Hubbard, D. D. (1992). Curriculum-based measurement and problem-solving assessment: Basic procedures and outcomes. Focus on Exceptional Children, 24, 1-20.

Shinn, M. R., Good, R. H., Knutson, N., Tilly, W. D., & Collins, V. L. (1992). Curriculum-based measurement of oral reading fluency: A confirmatory analysis of its relation to reading. School Psychology Review, 21, 459-479.

Slocum, T. A., Street, E. M., & Gilberts, G. (1995). A review of research and theory on the relationship between oral reading rate and reading comprehension. Journal of Behavioral Education, 5, 377-398.

Snow, C. E., Burns, M. S., & Griffin, P. (Eds.). (1998). Preventing reading difficulties in young children. Washington DC: National Academy Press.

Stanovich, K. E. (1980). Toward an interactive-compensatory model of individual differences in the development of reading fluency. Reading Research Quarterly, 1, 32-71.

Stanovich, K. E. (1981). Relationships between word decoding speed, general name-retrieval ability, and reading progress in first grade children. Journal of Educational Psychology, 73, 809-815.

Stanovich, K. E. (1986). Matthew effects in reading: Some consequences of individual differences in the acquisition of literacy. Reading Research Quarterly, 21, 360-406.

Stanovich, K. E., Cunningham, A. E., & West, R. F. (1981). A longitudinal study of the development of automatic recognition skills in first graders. Journal of Reading Behavior, 13, 57-74.

Tindal, G., Fuchs, L. S., Fuchs, D., Shinn, M. R., Deno, S. L., & Germann, G. (1983). The technical adequacy of a basal reading series mastery test: The Scott-Foresman reading program (Research Report No. 128). Minneapolis: University of Minnesota Institute for Research on Learning Disabilities.

Tindal, G., Fuchs, L. S., Fuchs, D., Shinn, M. R., Deno, S. L., & Germann, G. (1985). Empirical validation of criterion-referenced tests. Journal of Educational Research, 78, 203-209.

Tindal, G., Shinn, M. R., Fuchs, L. S., Fuchs, D., & Deno, S. L., Germann, G. (1983). The technical adequacy of a basal reading series mastery test (Research Report No. 113). Minneapolis: University of Minnesota Institute for Research on Learning Disabilities.

U.S. Department of Education. (2000). Twenty-second annual report to Congress on the implementation of the Individuals with Disabilities Education Act. Washington, DC : Author.

Wolf, M., Bally, H., & Morris, R. (1986). Automaticity, retrieval processes, and reading: A longitudinal study in average and impaired readers. Child Development, 57, 988-1000.

Woods, K. D., & Moe, A. J. (1985). Analytical Reading Inventory (3rd ed.). Columbus, OH: Merrill.

Table 1

Characteristics of Studies using Text Fluency

Study	Subjects	Passage Derived From	Criterion Measure	Correlations
Bain & Garlock, 1992	479 1 st to 3 rd grade (1 st was Title 1)	MacMillan Series r (at students grade level)	Comprehensive Tests of Basic Skills (CTBS)	.62 (1 st) .79 (2 nd) .72 (3 rd)
Buehner, 1983	92 4 th grade general ed.	Standard Reading Inventory & Scott-Foresman (2 nd grade level)	Comprehensive Tests of Basic Skills (CTBS)	.63
Collins, 1989	58 2 nd grade general ed.	Harcourt-Brace-Jovanovich (HBJ) Basal Reader (level 6 & 7)	California Achievement Test (CAT) & Harcourt- Brace-Jovanovich (HBJ) Basal Reader end of level test for comprehension	.75 .60
Deno, Mirkin, and Chiang, 1982; see also Fuchs, Tindal, and Deno, 1984.	45 1 st to 6 th grade with 18 LD	Allyn-Bacon, Ginn, & Houghton-Mifflin Basal Readers (3 rd & 6 th grade level)	Cloze Word Meaning	.86 (3 rd) .87 (6 th) .57 (3 rd) .56 (6 th)
Fuchs, 1981; see also Fuchs & Deno, 1992	91 1 st to 6 th grade with 15 special ed. & 23 Title 1	Ginn & Scott-Foresman Basal Readers, from each level (Ginn=10, S-F=9) (grade level 2 nd to 6 th)	Woodcock Reading Mastery Test (WRMT)	.91 (Ginn) .91 (S-F)

Fuchs, Fuchs, and Maxwell, 1988	35 4 th to 8 th grade with 27 LD, 7 ED, & 1 MR	Comprehensive Tests of Basic Skills (CRAB) at 2 nd & 3 rd grade level	SAT Questions Cloze Retell	.91 .84 .75 .74
Fuchs, Tindal, Fuchs et al., 1983	21 4 th grade general ed.	Word Reading Test (at students grade level)	Holt Basic Reading Series	.79
Fuchs, Tindal, Shinn, Fuchs, and Deno, 1983; see also Tindal, Fuchs et al., 1985	47 5 th grade general ed.	Word Reading Test (at students grade level)	Ginn Basal Reader Mastery Test	.72
Hintze, Shapiro, Conte, and Basile, 1997	57 2 nd to 4 th grade with 8 special ed.	Authentic & Literature based Basal Readers (1 st through 5 th grade level)	Degrees of Reading Power Test (DRP)	<u>.67 (Auth)</u>
Jenkins, Fuchs et al., 2000	113 4 th grade with 7 RD	Comprehensive Tests of Basic Skills (CRAB) at 2 nd & 3 rd grade level	Iowa Test of Basic Skills (ITBS)	.83
				.64 (Liter.)

Jenkins & Jewell, 1993	335 2 nd to 6 th grade with 15 LD, 1 MR, 1 ED	Narrative passages (1.7 grade level)	Gates-MacGinite Reading Test (GMRT)	.86(2 nd)
				.82(3 rd)
				.86(4 th)
				.68(5 th)
				.63(6 th)
			Metropolitan Achievement Test (MAT)	.84(2 nd)
	.67(3 rd)			
	.82(4 th)			
	.64(5 th)			
	.58(6 th)			
Kranzler, Brownell, and Miller, 1998	57 4 th grade general ed.	Ginn Basal Readers (4 th grade level)	Kaufman Test of Educational Achievement-Brief (KTEA -B)	.41
Kranzler, Miller, and Jordan, 1999	326 2 nd to 5 th grade general ed.	Ginn Basal Readers (2 nd through 5 th grade level)	California Achievement Test (CAT)	.63 (2 nd)
				.52 (3 rd)
				.54 (4 th)
	.51 (5 th)			
Lomax. 1983	101 1 st to 6 th grade LD	Diagnostic Reading Scales (DRS)	Diagnostic Reading Scales (DRS)	.76
			Comprehensive Tests of Basic Skills (CTBS)	.75

Madeline & Wheldall, 1998	50 1 st to 5 th grade general ed	Wheldall Assessment of Reading Passages, (5 th to 6 th grade reading level)	Neale Analysis of Reading- Revised	.71
Markell, 1991	42 3 rd grade general ed.	Holt, Rinehart, & Winston Basal Readers (2 nd , 4 th , & 6 th grade level)	Mazes below grade level	.86
			Mazes at grade level	.89
			Mazes above grade level	.87
			Questions below grade level	.26
			Questions at grade level	.50
			Questions below grade level	.40
Marston & Deno, 1982	26 3 rd grade general ed.	Ginn Reading Series (3 rd grade level)	Stanford Achievement Test (SAT)	.88
			Science Research Associates (SRA)	.80
			Ginn 720 Reading Series	.83
Parker, Hasbrouck, and Tindal, 1992	64 4 th to 6 th grade using lowest readers & Title 1	Students Basal Reader (1 below, 1 at and 1 above grade level)	Mazes	.76 (4 th)
				.76 (5 th)
				.59 (6 th)

Shinn, Good,	238	Harcourt-Brace-Jovanovich	Cloze	.77 (3 rd)
Knutson, Tilly, and Collins, 1992	3 rd & 5 th grades with 5% special ed.	(HBJ) Basal Reader (at students grade level)	SDRT	.63 (5 th) .59 (3 rd)
			Retell	.58 (5 th) .60 (3 rd) .39 (5 th)
Tindal, Fuchs et al., 1983; see also Tindal, Fuchs et al., 1985	25 4 th grade general ed.	Word Reading Test (at students grade level)	Scott-Foresman Basal Mastery Test	.70
Tindal, Shinn et al., 1983; see also Tindal, Fuchs et al., 1985	47 6 th grade general ed.	Word Reading Test (at students grade level)	Houghton-Mifflin Basic Reading Test	.66

Characteristics of Studies using List Fluency

Study	Subjects	Passage Derived From	Criterion Measure	Correlations
Deno, Mirkin, and	45	Basic Elementary Reading	Cloze	.84 (3 rd)
Chiang, 1982; see also Fuchs, Tindal, and Deno, 1984.	1 st to 6 th grade with 18 LD	Vocabulary-R series (3 rd & 6 th grade level)	Word Meaning	.86 (6 th) .60 (3 rd) .63 (6 th)
Fuchs, Tindal, Fuchs et al., 1983	21 4 th grade general ed.	Holt Basic Reading Series word list	Holt Basic Reading Series	.75
Fuchs, Tindal, Shinn, Fuchs, and Deno, 1983; see also Tindal, Fuchs et al., 1985	47 5 th grade general ed.	Ginn Basal Reader Mastery word list	Ginn Basal Reader Mastery Test	.69
Jenkins, Fuchs et al., 2000	113 4 th grade with 7 RD	Comprehensive Tests of Basic Skills (CRAB) word list	Iowa Test of Basic Skills (ITBS)	.53
Marston. 1982	83 3 rd to 6 th grade at-risk	Harris-Jacobson Word list	Stanford Achievement Test (SAT)	.72 (3 rd) .44 (at grade)

Tindal, Fuchs et al., 1983; see also Tindal, Fuchs et al., 1985	25 4 th grade general ed.	Scott-Foresman Basal Reader word list	Scott-Foresman Basal Mastery Test	.52
Tindal, Shinn et al., 1983; see also Tindal, Fuchs et al., 1985	47 6 th grade general ed.	Houghton-Mifflin Reader word list	Houghton-Mifflin Basic Reading Test	.55

Characteristics of Studies using Text Accuracy

Study	Subjects	Passage Derived From	Criterion Measure	Correlations
Fuchs. 1981	91 1 st to 6 th grade with 15 special ed. & 23 title 1	Ginn & Scott-Foresman Basal Readers	Woodcock Reading Mastery test (WRMT)	.82 (Ginn) .86 (Scott- Foresman)
Jenkins, Fuchs et al., 2000	113 4 th grade with 7 RD	Comprehensive Tests of Basic Skills (CRAB)	Iowa Test of Basic Skills (ITBS)	.53

Parker, Hasbrouck, and Tindal, 1992	64 4 th to 6 th grade lowest readers	Basal Reader Passages	Mazes	.58 (4 th) .30 (5 th) .71 (6 th)
--	--	-----------------------	-------	--

Characteristics of Studies using List Accuracy

Study	Subjects	Passage Derived From	Criterion Measure	Correlations
Bower and Swanson, 1991	46 2 nd grade 26 poor readers	Woodcock Reading Mastery Test (WRMT)	Canadian Test of Basic Skills	.50
Jenkins, Fuchs et al., 2000	113 4 th grade with 7 RD	Comprehensive Tests of Basic Skills (CRAB) word list	Iowa Test of Basic Skills (ITBS)	.61
Lomax, 1983	101 1 st to 6 th grade LD	Wide Range Achievement Test (WRAT) & Diagnostic Reading Scales (DRS)	Diagnostic Reading Scales (DRS) Comprehensive Tests of Basic Skills (CTBS)	.83 (WRAT) .89 (DRS) .77 (WRAT) .76 (DRS)
McCormick and Samuels, 1979	26 2 nd grade general ed.	Gray Oral Reading Test (GORT) & SRA Achievement Series – word list at 1 st and 2 nd grade level	Story retell and comprehension questions based on passages from the GORT & SRA Achievement Series.	.49(1 st) .70(2 nd)
Nation and Snowling, 1997	184 2 nd to 3 rd grade general ed.	Weschler Objective Reading Dimensions (WORD)- word list	Neale Analysis of Reading Ability- Revised Suffolk Reading Scale	.75 .89

Rice, 1981	106 1 st grade general ed.	Slosson Oral Reading Test	Cooperative Primary Test	.68
Shankweiler, Lundquist, et al. 1999	361 2 nd to 4 th grade 310 RD	Woodcock-Johnson Psycho- Educational Battery (WJ-R) and Wide Range Achievement Test- Revised (WRAT) combined word list	Woodcock-Johnson Psycho-Educational Battery (WJ- R), Gray Oral Reading Test (GORT), and Formal Reading Inventory all combined	.89

Table 2Mean Correlations for Text Fluency by Passage Level

Passage Level	Number of Correlations	<u>n</u>	<u>M</u>	<u>SD</u>	Range
<u>Below Grade</u>	8	575	.73	.10	.16 - .85
At Grade	12	998	.68	.11	.41 - .84

Table 3Mean Correlations for Text Fluency by Passage Length

Passage Length	Number of Correlations	<u>n</u>	<u>M</u>	<u>SD</u>	Range
90	1	66	.62	NA	NA
100	1	91	.91	NA	NA
172	1	42	.63	NA	NA
190	5	335	.72	.10	.61 - .85
200	6	247	.70	.04	.66 - .79
250	10	685	.57	.08	.41 - .76
300	3	129	.73	.06	.68 - .84
400	2	148	.83	.09	.81 - .83

Table 4Mean Correlations for Text Fluency by Graded or Common Passages

Passages	<u>2nd Grade</u>			<u>3rd Grade</u>			<u>4th Grade</u>			<u>5th Grade</u>			<u>6th Grade</u>		
	n	M	SD	n	M	SD	n	M	SD	n	M	SD	n	M	SD
<u>Below Grade Level</u>	<u>NA</u>	<u>NA</u>	<u>NA</u>	<u>NA</u>	<u>NA</u>	<u>NA</u>	<u>NA</u>	<u>NA</u>	<u>NA</u>	<u>NA</u>	<u>NA</u>	<u>NA</u>	<u>NA</u>	<u>NA</u>	<u>NA</u>
At Grade Level	256	.77	.04	241	.73	.04	103	.56	.17	171	.58	.09	47	.66	NA
Above Grade Level	<u>NA</u>	<u>NA</u>	<u>NA</u>	<u>NA</u>	<u>NA</u>	<u>NA</u>	<u>NA</u>	<u>NA</u>	<u>NA</u>	<u>NA</u>	<u>NA</u>	<u>NA</u>	<u>NA</u>	<u>NA</u>	<u>NA</u>
Common	47	.85	NA	50	.75	NA	271	.76	.10	47	.66	NA	125	.61	NA

Table 5Mean Correlations for Text Fluency by Criterion Measure

<u>Criterion Measure</u>	Number of Correlations	<u>n</u>	<u>M</u>	<u>SD</u>	Range
<u>Mazes</u>	5	212	.78	.12	.59 - .91
Basal Reader Test	4	140	.71	.04	.66 - .79
Questions on Passage Read	1	50	.71	NA	NA
Standardized, Group- Administered	15	1446	.69	.11	.51 - .85
Kaufman Test of Educational Achievement- Brief	1	57	.41	NA	NA

Table 6Mean Correlations for Text Fluency by Grade

<u>Grade</u>	Number of Correlations	<u>n</u>	<u>M</u>	<u>SD</u>	Range
Separate Grades					
1	1	66	.62	NA	NA
2	4	387	.75	.08	.63 - .85
3	6	523	.68	.08	.52 - .84
4	8	490	.68	.15	.41 - .84
5	5	311	.59	.09	.51 - .76
6	3	193	.62	.02	.59 - .66
Grouped Grades					
1 through 5	1	50	.71	NA	NA
1 through 6	3	237	.81	.08	.72 - .91
2 through 4	1	57	.66	NA	NA
4 through 8	1	35	.81	NA	NA

Table 7**Mean Correlations for Text Fluency by Student Group**

Student Group	Number of Correlations	<u>n</u>	<u>M</u>	<u>SD</u>	Range
<u>General Education</u>	16	1204	.66	.11	.41 - .84
Special Education	2	136	.77	.02	.76 - .81
Combination of General and Special Education	11	879	.71	.12	.53 - .91
At-Risk	4	130	.66	.07	.59 - .76

Table 8Mean Correlations for Text Fluency by Sample Size

Number of Correlations					
Sample Size		<u>n</u>	<u>M</u>	<u>SD</u>	Range
< 50	13	446	.73	.08	.59 - .85
50 - 99	13	913	.65	.13	.41 - .91
> 99	7	990	.70	.09	.53 - .83

Table 9Mean Correlations for List Fluency by Number of Words on List

Word List Length	Number of Correlations	<u>n</u>	<u>M</u>	<u>SD</u>	Range
60	1	452	.73	NA	NA
150	4	140	.62	.09	.52 - .75
400	1	113	.53	NA	NA

Table 10

Mean Correlations for Lists Fluency by List Level

Word List	Number of Correlations				
Level		<u>n</u>	<u>M</u>	<u>SD</u>	Range
<u>Below Grade</u>	1	113	.53	NA	NA
At Grade	4	140	.62	.09	.52 - .75

Table 11Mean Correlations for Lists Fluency by Criterion Measure

<u>Criterion Measure</u>	Number of				
	Correlations	<u>n</u>	<u>M</u>	<u>SD</u>	Range
Basal Reader Test	4	140	.62	.09	.52 - .75
Standardized Group-Administered	2	196	.55	.02	.53 - .58

Table 12Mean Correlations for Lists Fluency by Grade

<u>Grade</u>	Number of Correlations	<u>n</u>	<u>M</u>	<u>SD</u>	Range
Separate Grades					
4	3	159	.56	.08	.52 - .75
5	1	47	.69	NA	NA
6	1	47	.55	NA	NA
Grouped Grades					
1 through 6	1	45	.73	NA	NA
3 through 6	1	83	.58	NA	NA

Table 13Mean Correlations for Lists Fluency by Student Group

Student Group	Number of Correlations	<u>n</u>	<u>M</u>	<u>SD</u>	Range
<u>General Education</u>	4	140	.62	.09	.52 - .75
Combination of General and Special Education	2	158	.59	.09	.53 - .73
At-Risk	1	83	.58	NA	NA

Table 14Mean Correlations for Lists Fluency by Sample Size

Number of Correlations						
Sample Size		<u>n</u>	<u>M</u>	<u>SD</u>	Range	
< 50	5	185	.65	.09	.52 - .75	
50 - 99	1	83	.58	NA	NA	
> 99	1	113	.53	NA	NA	

Table 15Mean Correlations for Text Accuracy by Passage Length

Passage Length	Number of Correlations	<u>n</u>	<u>M</u>	<u>SD</u>	Range
100	1	91	.84	NA	NA
250	3	64	.53	.17	.30 - .71
400	1	113	.50	NA	NA

Table16**Mean Correlations for Text Accuracy by Criterion Measure**

<u>Criterion Measure</u>	Number of Correlations	<u>n</u>	<u>M</u>	<u>SD</u>	Range
<u>Mazes</u>	4	155	.71	.19	.30 - .84
Standardized Group-Administered	1	113	.50	NA	NA

Table 17Mean Correlations for Text Accuracy by Grade

<u>Grade</u>	Number of Correlations	<u>n</u>	<u>M</u>	<u>SD</u>	Range
Separate Grades					
4	2	135	.51	.03	.50 - .58
5	1	21	.30	NA	NA
6	1	21	.71	NA	NA
Grouped Grades					
1 through 6	1	91	.84	NA	NA

Table 18

Mean Correlations for Text Accuracy by Student Group

Student Group	Number of Correlations	n	<u>M</u>	<u>SD</u>	Range
Combination of General and Special Education	2	204	.65	.17	.50 - .84
At-Risk	3	64	.53	.17	.30 -.71

Table 19Mean Correlations for Text Accuracy by Sample Size

Number of Correlations					
Sample Size		<u>n</u>	<u>M</u>	<u>SD</u>	Range
< 50	3	64	.53	.17	.30 - .71
50 - 99	1	91	.84	NA	NA
> 99	1	113	.50	NA	NA

Table 20Mean Correlations for List Accuracy by List Length

List Length	Number of Correlations	<u>n</u>	<u>M</u>	<u>SD</u>	Range
20	1	106	.68	NA	NA
400	1	113	.61	NA	NA

Table 21**Mean Correlations for List Accuracy by Criterion Measure**

<u>Criterion Measure</u>	Number of Correlations	<u>n</u>	<u>M</u>	<u>SD</u>	Range
Woodcock-Johnson Psycho- educational Battery - Revised	1	361	.89	NA	NA
Standardized Group-Administered	4	366	.67	.10	.50 - .81
Story Retell with Questions	1	26	.60	NA	NA

Table 22Mean Correlations for List Accuracy by Grade

<u>Grade</u>	Number of Correlations	<u>n</u>	<u>M</u>	<u>SD</u>	Range
Separate Grades					
1	1	106	.68	NA	NA
2	2	72	.54	.05	.50 - .60
4	1	113	.61	NA	NA
Grouped Grades					
1 through 6	1	101	.81	NA	NA
2 through 3	1	184	.82	NA	NA
2 through 4	1	361	.89	NA	NA

Table 23**Mean Correlations for List Accuracy by Student Group**

Student Group	Number of Correlations	<u>n</u>	<u>M</u>	<u>SD</u>	Range
<u>General Education</u>	3	316	.75	.08	.60 - .82
<u>Special Education</u>	1	101	.81	NA	NA
Combination of General and Special Education	3	520	.79	.15	.50 - .89

Table 24Mean Correlations for List Accuracy by Sample Size

Sample Size	Number of Correlations	<u>n</u>	<u>M</u>	<u>SD</u>	Range
<u>< 50</u>	2	72	.54	.05	.50 - .60
> 99	5	865	.80	.10	.61 - .89