

# Reading & Writing Quarterly

## Overcoming Learning Difficulties

ISSN: (Print) (Online) Journal homepage: <https://www.tandfonline.com/loi/urwl20>

## The Validity of Two Tests of Silent Reading Fluency: A Meta-Analytic Review

Daniel R. Wissinger, Adrea J. Truckenmiller, Amber E. Konek & Stephen Ciullo

**To cite this article:** Daniel R. Wissinger, Adrea J. Truckenmiller, Amber E. Konek & Stephen Ciullo (2023): The Validity of Two Tests of Silent Reading Fluency: A Meta-Analytic Review, *Reading & Writing Quarterly*, DOI: [10.1080/10573569.2023.2175340](https://doi.org/10.1080/10573569.2023.2175340)

**To link to this article:** <https://doi.org/10.1080/10573569.2023.2175340>



© 2023 The Author(s). Published with license by Taylor & Francis Group, LLC.



Published online: 19 Mar 2023.



Submit your article to this journal [↗](#)



Article views: 671





View related articles [↗](#)



View Crossmark data [↗](#)

# The Validity of Two Tests of Silent Reading Fluency: A Meta-Analytic Review


Daniel R. Wissinger<sup>a</sup> , Adrea J. Truckenmiller<sup>b</sup> , Amber E. Konek<sup>a</sup>, and Stephen Ciullo<sup>c</sup>

<sup>a</sup>Indiana University of Pennsylvania, Indiana, PA, USA; <sup>b</sup>Michigan State University, East Lansing, MI, USA; <sup>c</sup>Department of Curriculum and Instruction, Texas State University

## ABSTRACT

The purpose of this meta-analysis was to evaluate the potential of two silent reading fluency measures as indicators of reading competence. Specifically, we analyzed score differences between the *Test of Silent Contextual Reading Fluency* (TOSCRF), the *Test of Silent Word Reading Fluency* (TOSWRF), and other standardized measures of reading to determine whether the two silent reading fluency measures were valid indicators of reading competence. Further, potential moderating variables were examined: (a) type of criterion reading measure (i.e., decoding/encoding, word-letter identification, fluency, vocabulary, comprehension); (b) type of silent reading fluency measure (i.e., word vs. contextual); (c) type of learner (English language learner [ELL] status, at risk for a disability; average; above average and gifted); and (e) administration format (i.e., group or individual), and the reading score outcomes. A comparison of effect sizes, across 47 studies and 47,616 participants, revealed the very little score differences between the TOSCRF, TOSWRF, and other standardized measures of reading competence ( $r = 0.07$ , very small or trivial). Three moderator variables (English language learner status, type of silent reading fluency measure [word vs. contextual], and administration format [individual vs. group]) did, in fact, moderate effect sizes across studies. A discussion of the implications for using the TOSCRF and TOSWRF as indicators of reading competence, study limitations, and recommendations for future research are included.

For years, oral reading fluency measures have been widely used by teachers and school districts to screen and progress monitor students' reading development. While oral reading fluency has been established as an important indicator of reading comprehension and overall reading competence (Fuchs, 2004; Kara et al., 2020; Reschly et al., 2009), silent reading fluency has been relatively unexplored, mostly due to a lack of validated, norm-referenced measures. Two commonly used measures of silent reading fluency in the extant research are the *Test of Silent Reading Word Fluency—Second Edition* (TOSWRF-2; Mather et al., 2014), and its companion, the *Test of Silent Contextual Reading Fluency—Second Edition* (TOSCRF-2; Hammill et al., 2014). Although these tests have been used by researchers and school professionals for almost two decades, no extensive review of their associations with aspects of reading performance has been published. This paper

**CONTACT** Daniel R. Wissinger  [Daniel.Wissinger@iup.edu](mailto:Daniel.Wissinger@iup.edu)  College of Education and Communications, Professional Studies in Education Department, Indiana University of Pennsylvania, 312 Davis Hall, 570 South Eleventh Street, Indiana, PA 15705-1080, USA.

© 2023 The Author(s). Published with license by Taylor & Francis Group, LLC.

This is an Open Access article distributed under the terms of the Creative Commons Attribution-NonCommercial-NoDerivatives License (<http://creativecommons.org/licenses/by-nc-nd/4.0/>), which permits non-commercial re-use, distribution, and reproduction in any medium, provided the original work is properly cited, and is not altered, transformed, or built upon in any way. The terms on which this article has been published allow the posting of the Accepted Manuscript in a repository by the author(s) or with their consent.

begins with a description of the theoretical role of silent reading fluency, followed by a description of the TOSWRF-2 and TOSCRF-2 and concludes with a practice-oriented rationale for synthesizing evidence of validity for these tests.

### ***Silent reading fluency***

Fluent reading has been defined as accurate, effortless, and automatic word recognition that is facilitated by reading comprehension (Reutzel & Juth, 2014). Fluent reading can occur in the oral mode where students read aloud with appropriate phrasing or in the silent mode where students process text automatically but do not include oral-motor output. However, oral and silent reading fluency are not the same (Denton et al., 2011; Price et al., 2016). In empirical models, while silent reading fluency is highly correlated with latent constructs of decoding, oral reading fluency, and comprehension, research indicates that silent reading fluency does not directly measure decoding, fluency, or comprehension (Cirino et al., 2013). Rather, it is a general outcome measure (Espin & Deno, 2016).

The very high correlations make silent reading fluency an efficient general reading outcome metric in a practical context. For example, schools need a general outcome reading measure that can be efficiently administered more than once per year to monitor reading instruction effectiveness for their Tier 1 general instruction, for students in supplemental or intensive intervention, and students receiving special education services. Typically, practitioners and researchers need to rely on achievement tests to measure instructional impact. Many of these achievement tests need to be individually administered, require specialized training to administer, require a large amount of testing time for each student (e.g., longer than 10 minutes), and cannot be administered more than twice per year. Oral reading fluency measures require very little testing time (i.e., less than 5 minutes) and can be administered repeatedly for the purpose of monitoring progress in response to instruction (Fuchs et al., 2001). It is also a general outcome metric associated with multiple aspects of reading, such as decoding and reading comprehension (Fuchs et al., 2001). However, in later grade levels, students have higher silent reading fluency than oral reading fluency (van den Boer et al., 2022). Therefore, it is important to analyze the function of silent reading fluency as an efficient and valid way to measure progress. It is also important to determine if silent reading fluency measures provide similar results as the achievement measures of different reading domains.

The reading domains commonly measured in reading achievement batteries (e.g., Gates MacGinitie, Wechsler Individual Achievement Test, Woodcock Johnson Tests of Achievement) include decoding, fluency, letter-word identification, vocabulary, and comprehension. If silent reading fluency measures provide similar results as measures of these domains, silent reading fluency may be a more efficient alternative to monitoring student progress in schools and research. In the current meta-analysis, we aim to evaluate the comparability of scores on silent reading fluency measures to achievement tests in a variety of domains.

Importantly, there are several measurement methods used to measure silent fluency, including sentence verification tasks (e.g., Test of Silent Reading Efficiency and Comprehension), underlining for comprehending paragraphs (Price et al., 2012), maze, comprehension-based silent reading rate (Hiebert et al., 2012), and “slasher” methods (e.g., TOSWRF and TOSCRF). These methods have similar but slightly different rates of classification accuracy as measures of general outcome measures (see Denton et al., 2011). The current meta-analysis focuses on aggregating the capacity of the slasher method to measure various aspects of reading.

### ***Description of the TOSWRF-2 and TOSCRF-2***

The TOSWRF and the TOSCRF were originally published in 2004 and 2006, respectively. In 2014, the second edition of both tests were published. The updated editions expanded the test age

range to include college-aged students, updated instructions, and added specific instructions for deaf and hearing-impaired students; however, the general procedures for administering and scoring are the same as those in the original version (Hammill et al., 2014; Mather et al., 2014). Therefore, in this analysis, we simply refer to original and updated editions as the TOSCRF and TOSWRF. This section will briefly review each of the tests.

### **The test of silent word reading fluency**

The TOSWRF is a word chaining test measures word identification, word comprehension, and reading speed in individuals between the ages of 6 years 3 months and 24 years 11 months. TOSWRF is primarily a measure of word identification, word comprehension, and reading speed. The test includes four equivalent forms (Forms A, B, C, and D) consisting of 220 unrelated words printed in rows with no spaces between them. The words are printed in lowercase and begin with pre-primer level words and increase in difficulty to adult-level words. The words were selected to have no words within them and such that no new words are create across the words when they were placed beside each other. The students are asked to draw a line between the boundaries of as many recognizable words as possible within three minutes (e.g., dimhowbluefig would result in a divided chain of dim/how/blue/fig).

The test can be administered individually or to an entire classroom of students in three-minutes and yields raw scores (based on total words correctly identified), standard scores ( $M = 100$ ,  $SD = 15$ ), percentiles, and age and grade equivalents. The norms are based on a representative sample ( $N = 2,429$ ) ranging in age from 6–13 to 24–11 years in 35 states. Studies with the TOSWRF show it is valid and reliable for a wide variety of subgroups and the general population (e.g., Mather et al., 2004). Test-retest reliability estimates of .92 and .91 are reported for Forms A and B, respectively. Alternate-form reliability for immediate and delayed is reported as .86 and .89, respectively. Correlations between the TOSWRF and other standardized measures (e.g., *The Comprehensive Test of Basic Skills* [CTBS]; *Woodcock-Johnson III* [WJ-III]) were moderate to large (range 0.57 to 0.77).

### **The test of silent contextual reading fluency**

The TOSCRF can also be administered to groups or individuals. Unlike the TOSWRF, however, it assesses contextual reading abilities (i.e., word identification, vocabulary, sentence structure, comprehension, and fluency) in students ranging in age from 7 years 0 months to 24 years 11 months. It has four forms (Forms A, B, C, and D). Each form consists of a series of 12 passages. The words of each passage are printed in uppercase, omitting spaces and punctuation (e.g., AYELLOWBIRDWITHBLUEWINGS). Students are provided three-minutes to draw lines between as all the words in as many sentences as they can. The raw score is based on the total number of words correctly identified. This format allows the examiner to document how many words a student can recognize without having to rely on the student's oral production of words (Hammill et al., 2014). This format also provides a practical advantage in terms of assessment administration because it can be administered to a group of students in just three minutes compared with the labor-intensive oral reading probes which requires administrators to listen to each child one at a time.

The TOSCRF yields raw scores, standard scores ( $M = 100$ ,  $SD = 15$ ), percentiles, and age and grade equivalents. This measure was normed using a national representative sample of 2,375 individuals in 29 states. Alternative form, test-retest, and alternate form coefficients are roughly similar to oral reading fluency reliability (range = 0.82–0.93). The Examiner's Manual provides strong validity for the test as a measure of reading ability. Average correlations between the TOSCRF and other criterion reading tests (e.g., *Stanford Achievement Test Series – Ninth Edition*,

*Test of Word Reading Efficiency* [TOWRE], and *Woodcock-Johnson III* [WJ-III]) were large to very large (range 0.67 to 0.85).

### ***Purpose and procedures of the measurements***

The TOSCRF and TOSWRF are measures that can be administered to individuals or groups in three-minutes. Like tests of oral reading fluency, they do not attempt to measure reading subskills (e.g., decoding, letter-word identification, vocabulary, and comprehension) or broad-spectrum correlate skills that, often, are difficult to use in classroom settings because they require one-on-one assessment sessions, specially trained personnel, and considerable time commitments (Kamhi & Catts, 2017). Rather, the TOSCRF and TOSWRF provide an indication of a student's general reading achievement level.

Like oral reading fluency scores, silent reading fluency scores reflect more than just reading fluency (e.g., Denton et al., 2011). The authors of the TOSCRF and TOSWRF found that students who perform better on silent reading fluency tasks also perform better on measures of vocabulary, sentence structure, comprehension, and oral reading fluency. In their manuals, Hammill et al. (2014) and Mather et al. (2014) suggest that performance on their brief silent reading fluency measure actually reflects general reading performance and can be useful as a progress monitoring measure. As such, the TOSCRF and TOSWRF have the potential to identify good and poor readers and to be used for screening, progress monitoring, or research purposes. The current study seeks meta-analytic evidence of the reading domains for which the TOSWRF and TOSCRF may serve as general outcome measures.

### ***Purpose of the present study***

Growing evidence suggests that silent reading fluency is a valid indicator of students' overall reading competence (Kim et al., 2011; Rasinski et al., 2011). It is, therefore, important to document the validity of silent reading measures for use in identifying good and poor readers and monitoring progress in interventions. Despite the fact that the original versions of the TOSWRF and TOSCRF have been in print for well over a decade (Hammill et al., 2006; Mather et al., 2004), to our knowledge, previous systematic reviews of literature have yet to synthesize validity studies of these tests into a meta-analysis. Examining the relationship between the TOSWRF, TOSCRF, and a weighted average of the two tests and a variety of instructionally relevant reading skills (decoding/encoding, word-letter identification, fluency, vocabulary, and comprehension) was an important goal of this study to determine if the TOSCRF and TOSWRF under- or over-estimated achievement in any of the specific skills. The primary purpose of this meta-analysis, therefore, was to examine the extent to which these silent reading measures are valid indicators of reading competence by examining experimental effect sizes and standardized mean differences ( $r$ ) between the TOSWRF, TOSCRF, the weighted average of the two silent reading measures, and other norm-referenced, standardized tests of reading. When the mean scale scores are equivalent across tests (i.e., do not differ), the mean experimental effect size would be equal to 0.00. It was hypothesized that differences between the TOSWRF, TOSCRF, and their weighted average, and other standardized tests of reading would be small to trivial.

The secondary purpose of this meta-analysis was to investigate variables that have been shown to moderate students' performance on standardized reading measures in previous research (Abedi, 2002; Connor et al., 2014; Shin & McMaster, 2019). Specifically, we examined the relationship between performance and: (a) domain of the criterion reading measure (i.e., decoding/encoding, word-letter identification, fluency, vocabulary, comprehension); (b) type of learner (English language learner [ELL] status, at risk for a disability; average; above average

and gifted); and (c) administration format (i.e., group or individual administration, word vs. contextual format).

Based on the rationale and purposes provided above, the two following research questions guided this study:

1. **Research question 1:** Do scores from the TOSWRF, the TOSCRF, and a weighted average of these two silent reading measures differ from those from other standardized measures of reading competence?
2. **Research question 2:** What variables moderate the relationship between the TOSWRF, TOSCRF, and other standardized measures of reading competence?

## Method

We conducted the methodological portion of this study in three stages, using procedures adapted from previous meta-analytic studies that examined relationships between standardized measures of reading achievement (e.g., Reschly et al., 2009). The stages included the following: (a) literature search and study selection, (b) data coding and reliability procedures, and (c) effect size calculation and data analysis.

### Stage 1: Literature search and study selection

Stage one of this meta-analysis involved an extensive literature search and resulting study selection. Important procedures used in the literature search and subsequent study selection are described here.

#### Literature search

We applied a series of strategies to identify studies for this review. First, we selected key terms from language commonly used in the literature and entered in the following databases: ERIC, EBSCOHost, Education Source, Google Scholar, ProQuest Dissertations & Theses, PsycArticles, PsycINFO, Google Scholar, and WEB of Science. Descriptors for the searches included combinations of the following keywords: silent reading fluency\*, silent reading rate\*, comprehension-based silent reading\*, measurement\*, assessment\*, tests\*, and progress monitoring\* with no restriction on where the terms occur (i.e., title, abstract, descriptor, or full text). We also included norm-referenced\* OR standardized\* OR aptitude\* OR competence\* in the search string to focus on studies that have used standardized reading measures and report subtest and/or composite scores. Next, researchers conducted a hand search of journals that frequently publish research in this domain, including the following: *Assessment for Effective Intervention*, *Reading Research Quarterly*, *Review of Educational Research*, *Exceptional Children*, *Journal of Special Education*, *Journal of Learning Disabilities*, *Remedial and Special Education*, *Journal of School Psychology*, *School Psychology Quarterly*, and *School Psychology Review*. Third, we examined references and abstracts from a list of authors who commonly publish on the topic of silent reading fluency (e.g., Berendes et al., 2019; Denton et al., 2011; Freeland et al., 2000; Kim et al., 2011; Price et al., 2012; 2016; Rasinski et al., 2016). Fourth, we contacted the authors of both the TOSCRF and TOSWRF and asked them to share any relevant papers or citations with us, including ones that were conducted by themselves or others. Fifth, we obtained the technical manuals for both the TOSCRF and TOSWRF, which provided descriptive statistics, correlations, and other data sources examining relationships between the TOSWRF, TOSCRF, and other standardized assessment scores. Last, when the full-text of a document was obtained from one of the search strategies, we searched the reference list to identify other potential studies.

### **Study selection**

The search yielded 5,234 possible articles; 1,112 were duplicates. We screened the title and abstracts of each of the remaining 4,122 articles to identify possible studies that met the inclusion/exclusion criteria, which resulted in 145 articles. These articles were reviewed in full by the first author and a trained graduate research assistant. Studies were eliminated for the following reasons: (a) did not report scores for the TOSCRF and/or TOSWRF, and other norm-referenced, standardized tests of reading competence, (b) missing statistical data needed to calculate effects, (c) the TOSWRF and/or TOSCRF were not administered and scored according to standard criteria, or prior to or within the same academic year as other norm-referenced, standardized tests of reading competence, (d) were not administered to individuals between 6 and 24 years of age to align with sampling norms, and (e) if standardized achievement scores were used to compare performance over multiple academic years or if assessment tasks were modified for use with special populations of students (as was done with R-CBM measures in Allinder & Eccarius, 1999). Based on these criteria, 47 studies were retained and included in the meta-analysis.

### **Stage 2: Data coding and reliability procedures**

Stage two of this meta-analysis involved standardized coding of the data and specific coding criteria and reliability checks to ensure fidelity across the coders. Data coding and reliability procedures used in this meta-analysis are described in this section.

### **Data coding**

The data coding and accompanying protocols were completed by the first and third author and guided by procedures used in previous meta-analysis (e.g., Finger & Ones, 1999; Reschly et al., 2009). First, general data from each study were coded. This included study information (i.e., author, date of publication) and sample demographics (i.e., ethnicity, age, gender, grade-level). Next, data relating to reading competence tests were coded (i.e., mean scores, standard deviations, and sample sizes). Last, categorical data prior research suggested might moderate effect sizes between tests of reading competence were extracted from the 47 studies and/or reports and are described below.

**Type of reading skill assessed by the criterion reading measure.** Mean standard scores and standard deviations from the 47 studies were coded according to the type of standard score reported (subtest or composite score) and type of reading skill measured by the criterion test (e.g., decoding, word-letter identification, fluency, vocabulary, and comprehension) based on descriptions provided by the authors in each study. If the criterion test or score *primarily* measured one of five component reading skills (e.g., decoding, word-letter identification, fluency, vocabulary, and comprehension), it was coded in the category described by the study. If the measure or score primarily captured more than two component reading skills, it was placed in the composite scores category. No measures or scores were excluded because each study included a rationale for including the measure. We conceptualize the composite scores category as a general achievement construct. A list of reading tests and their assigned categories can be viewed in [Table 1](#).

**Type of silent reading fluency measure.** The variable for type of silent reading fluency measure (i.e., word or contextual) required only a single variable. A dichotomous variable was created to categorize effect sizes by the two types of tests of silent reading fluency: silent contextual and silent word reading.



**Table 1.** Tests, subtests, and administration type.

Comprehension		Decoding/encoding	
GRADE reading comprehension	I	WJ-III Word Attack	I
WRMT-R Passage Comprehension	I	TOWRE Decoding	I
GORT-5 Comprehension	I	WIAT-III Pseudoword Decoding	I
WIAT-II Reading Comprehension	I	WJ-III Spelling	I
TOSREC Reading Comprehension	G	WIAT-III Spelling	I
<b>Fluency</b>		KTEA-II Spelling	I
DIBELS ORF	I	KTEA-II Nonsense Word Decoding	I
GORT-5 Rate	I	KTEA-II Decoding Composite	I
GORT-5 Accuracy	I	KTEA-II Decoding Fluency	I
GORT-5 Oral Reading Index	I	WRAT-3 Spelling	I
TOSCRF Fluency	G	CTBS Spelling	I
WJ-III Writing Sentence Composition	I	<b>Composite scores</b>	
KTEA-II Reading Fluency	I	ERA Reading Index	I
WJ III Reading Fluency	I	TOWRE	I
<b>Vocabulary</b>		GMRT	I
WIAT-III Picture Vocabulary	I	GORT-5	I
WIAT-III Rapid Picture Naming	I	KTEA-II	I
Peabody Picture Vocabulary Test-4	I	NDRT General reading ability	G
Peabody Picture Vocabulary Test-4	I	TERA-4 General reading index	I
<b>Word and letter identification</b>		KBIT nonverbal	I
KTEA-II Letter and Word Recognition	I	Test of Orthographic Competence	I
Rapid Automatic Naming	I	KTEA-III Reading	I
WJ-III Letter Word Identification	I	WIST Fundamental Literacy Index	I
TOWRE Sight Word Efficacy	I	TWS-5 Standard Score Form A	G
WIAT-III Word Reading	I	WRAT-3 Reading	I
WJ-III Word Identification	I	CTBS Reading	I
TOWRE Word Reading Efficiency	I	WJ III Broad Reading	I
WIAT-2 Word Reading Accuracy	I		
WIAT-2 Nonword Reading Accuracy	I		
KTEA-II Word Recognition Fluency	I		

*Note.* CTBS = Comprehensive Test of Basic Skills; DIBELS = Dynamic Indicators of Basic Early Literacy Skills; ERA = Early Reading Assessment; G = group administered; GORT = Gray Oral Reading Fluency; GRADE = Group Reading Assessment and Diagnostic Evaluation; I = individually administered; KBIT = Kaufman Brief Intelligence Test; KTEA = Kaufman Test of Educational Achievement; NDRT = Nelson-Denny Reading Test; RAN = Rapid Automatic Naming; SAT = Stanford Achievement Test; TERA = Test of Early Reading Ability; TOAL = Test of Adolescent and Adult Language; TOC = Test of Orthographic Competence; TOSCRF = Test of Silent Contextual Reading Fluency; TOSREC = Test of Silent Reading Efficacy and Comprehension; TOSWRF = Test of Silent Word Reading Fluency; TWS = Test of Written Spelling; WIAT = Wechsler Individual Achievement Test; WIST = Word Identification and Spelling Test; WJ = Woodcock-Johnson; WRAT = Wide Range Reading Achievement Test; WRMT-R = Woodcock Reading Mastery Test-Revised.

**Type of learner.** The type of learner variable required four codes. Students in all 47 studies were organized as follows: (a) ELL, (b) students with or at risk for disabilities (c) average learners; and (d) above average and gifted learners.

**Administration format.** The variable for administration format required two codes. Each test was coded according to how it was administered—either individually or to groups of students.

**Reliability procedures**

The first author developed a coding manual that included definitions of terms and examples and non-examples in each category, then followed standard reliability training and procedures (e.g., Reschly et al., 2009). Working with the third author, they coded a series of studies before assessing and establishing interrater reliability (benchmark was greater than 90%). After the 90% benchmark was achieved, each study was double coded independently by the first and third author. Finally, after each study was double coded, each code sheet was compared to identify potential discrepancies. Coding discrepancies were highlighted, discussed, and resolved through dialogue. Total interrater coding reliability was 93%.



### **Stage 3: Effect size calculation and data analysis**

Stage 3 of this meta-analysis involved calculating effect sizes for each of the reported TOSWRF, TOSCRF, and criterion reading measures. Once each effect size was calculated, the data were analyzed to answer our two research questions. This section described the procedures we used at this stage.

#### **Effect size calculation**

We summarized descriptive statistics collected from the 47 studies. Using Comprehensive Meta-Analysis (CMA) software (version 3.0; Biostat), weighted mean scores were calculated across the TOSWRF, TOSCRF, the subtests of component reading skills, and composite reading scores. Standardized mean difference effect sizes were calculated to examine the degree to which the TOSWRF and TOSCRF differed from the other standardized measures of reading. If a study did not report means and standard deviations, group comparison statistics (e.g., *F* test, *t*-test) were entered into CMA software to obtain an effect size.

#### **Average weighted effects across studies**

A total of 139 effect sizes were calculated from the 47 studies included in the analysis. We used a Fisher's *r*-to-*Z* transformation as the effect size metric. The formula is as follows:  $z_r = \ln((1+r)/(1-r))/2$  with a sample variance of  $1/\sqrt{N-3}$ , where *N* is the sample size associated with the respective bivariate correlation. All *Z* values were later transformed back to the *r* metric for ease of reporting. Because most of the studies reported more than one effect size, there was within-study dependence among effect sizes; therefore, a robust variance estimation (RVE; Hedges et al., 2010) approach was used to adjust standard errors to account for correlations among dependent effect sizes. Using the correlated effects method within the ROBUMETA package (see Tanner-Smith & Tipton, 2014, for details) we assumed a general between-outcomes correlation of  $\rho = .80$ . Two measures of variability were also computed: *Q* and *I*<sup>2</sup>. The *Q* statistic determines if variability in an average weighted effect size exceeds sampling error alone (Shadish & Haddock, 2009). *I*<sup>2</sup> was also calculated as a second complementary measure of homogeneity, as it is less sensitive to sample size than *Q* (Higgins & Green, 2011).

Extreme outliers were defined using Turney's (1977) definition of 1.5 interquartile ranges beyond 75th percentile upper and 25th percentile lower boundaries for effect sizes. To identify extreme outliers, the interquartile range was calculated and multiplied by 1.5 and effect sizes outside these 1.5 interquartile boundaries were identified. Next, we conducted a sensitivity analysis by estimating the overall mean effect size with identified outliers excluded. Sensitivity analysis results were compared to the overall mean effect size. This determined if the outliers were placing undue influence on the overall effect size. Last, we visually assessed funnel plots and employed Egger's test (Egger et al., 1997) to explore the possibility of publication bias.

When homogeneity in effect sizes exceeded sampling error alone (as represented by a statistically significant *Q* statistic), moderator analyses were conducted to determine if the excess variability could be accounted for by identifiable differences between studies. The moderators of interest included: (a) type of criterion reading measure (i.e., decoding/encoding, word-letter identification, fluency, vocabulary, comprehension); (b) type of silent reading fluency measure (i.e., word vs. contextual); (c) type of learner (English language learner [ELL] status, at risk for a disability; average; above average and gifted); and (e) administration format (i.e., group or individual).

## Results

### Research Question 1

*Do scores from the TOSWRF, the TOSCRF, and a weighted average of these two silent reading measures differ from those from other standardized measures of reading competence?*

Table 2 provides an overview of the characteristics of the studies included in this meta-analysis as well as the descriptive statistics for each individual effect size. The reading tests included in this analysis were administered to 47,616 children, adolescents, or young adults. A total of 139 mean scores from participants were extracted from 47 studies. Fifty-nine percent of the participants ( $n = 82$ ) were considered average learners; the remaining participants were students with or at-risk for disabilities ( $n = 49$ ), English language learners ( $n = 5$ ), or above average and gifted learners ( $n = 1$ ).

Twenty-eight of the 47 studies in the analysis reported outcomes on the TOSWRF; the remaining 19 studies reported outcomes on the TOSCRF. Because scores from both tests were reported using standardized scores (i.e.,  $M = 100$ ;  $SD = 15$ ), descriptive statistics were combined into a weighted mean score for each test, then combined into a comprehensive weighted average using CMA software (version 3.0; Biostat).

Table 2 reports the descriptive statistics, weighted means and standard deviations (with 95% confidence interval), and the correlation coefficient effect size for each of the measures of reading. The weighted mean score across the 75 reported scores on the TOSWRF was 96.83 ( $SD = 8.43$ ). The weighted mean score across the 65 reported scores on the TOSCRF was 91.56 ( $SD = 8.52$ ).

In the 47 studies examined, there were also 48 separate standardized tests of reading competence. Thirty-one of these tests were subtests of component reading skills and 16 (or 33% of the sample) were composite scores. The subtests were as follows: decoding ( $n = 13$ ); word-letter identification ( $n = 10$ ); fluency ( $n = 9$ ); vocabulary ( $n = 3$ ); and comprehension ( $n = 6$ ). Weighted mean scores were calculated for each subtest and composite score, then combined into a total weighted mean score (see Table 2). Weighted mean standard scores (with standard deviations in parentheses) were 97.74 (6.57) for decoding, 95.14 (9.57) for word-letter identification, 91.14 (14.40) for fluency, 96.25 (1.95) for vocabulary, 95.90 (6.81) for comprehension, and 99.60 (5.97) for composite scores. The weighted mean standard scores suggested that when the two tests of silent reading fluency and other measures of reading (i.e., component skills and composite scores) were administered to the same sample of participants, the difference between the standard scores is, on average, 2 standard score points (range of standard score differences = 0.61 to 5.07;  $M = 1.52$ ).

To determine the meaningfulness of these weighted mean score differences, we calculated the mean effect size difference between the TOSWRF, TOSCRF, and the criterion tests of reading ability (see the right three columns in Table 2). In interpreting the magnitude of effect sizes, we are guided by Hopkins (2002). He suggested that effect size  $r$  coefficients between .00 and .09 are *very small* or *trivial*, coefficients between .10 and .29 are *small*, coefficients between .30 and .49 are *moderate*, coefficients between .50 and .69 are *large*, coefficients between .70 and .89 are *very large*, and coefficients between .90 and 1.00 are *nearly perfect*. The median average effect size difference between the TOSWRF and the tests of reading competence was .03 (very small or trivial) and ranged from  $-.16$  (small) to .23 (small). The median average effect size difference between the TOSCRF and the tests of reading competence .16 (small) and ranged from  $-.44$  (moderate) to .12 (small). These results indicate that, on average, a student's performance on the TOSCRF and TOSWRF will mirror performance on other norm-referenced tests reading skills (and vice-versa).

When results from TOSWRF and TOSCRF were combined into a weighted mean score and compared to the weighted mean scores from the other measures of reading competence we found the scores were similarly nearly identical. The  $r$  value and magnitude in the last two columns in the bottom row of Table 2 represents the average weighted standard deviation unit test difference and qualitative difference between the TOSCRF and TOSWRF and the combined results from

**Table 2.** Descriptive Statistics, Weighted Mean Scores, and Effect Sizes for the TOSCRF, TOSWRF and Standardized Tests of Reading Competence.

Test	N	Mean (SD)	95% CI	TOSWRF		TOSCRF		Combined	
				r	magnitude	r	magnitude	r	magnitude
Silent reading fluency									
TOSWRF	75	96.83 (8.43)	[94.89, 98.77]	.00	***	.29	Small	-.13	Small
TOSCRF	58	91.56 (8.52)	[89.32, 93.80]	-.29	Small	.00	***	.15	Small
Weighted mean score	133	94.53 (8.84)	[93.02, 96.05]	-.13	Small	.15	Small	.00	***
Tests of reading competence									
Subtests									
Decoding	22	97.74 (6.57)	[94.25, 101.22]	-.05	Trivial	-.32	Moderate	-.13	Small
Word-Letter ID	37	95.14 (9.57)	[89.95, 100.32]	.09	Trivial	-.19	Small	-.03	Trivial
Fluency	22	91.14 (14.40)	[84.76, 97.52]	.23	Small	.02	Trivial	.12	Small
Vocabulary	3	96.25 (1.95)	[84.76, 97.53]	.01	Trivial	-.12	Small	-.03	Trivial
Comprehension	16	95.90 (6.81)	[92.27, 99.54]	.04	Trivial	.12	Small	-.05	Trivial
Composite scores	30	99.60 (5.97)	[96.62, 102.57]	-.16	Small	-.44	Moderate	-.23	Small
Weighted mean score	130	96.05 (11.78)	[94.01, 98.10]	.04	Trivial	-.18	Small	-.07	Trivial

Note. SD = standard deviation; CI = confidence intervals; TOSCRF = Test of Silent Contextual Reading Fluency; TOSWRF = Test of Silent Word Reading Fluency; r = correlation coefficient effect size; Magnitude = effect size differences based on Hopkins (2002).

other norm-referenced standardized measures of reading competence. The weighted mean average across the TOSWRF and TOSCRF was 94.53 ( $SD=8.84$ ) and the weighted mean score across the 48 standardized tests of reading competence was 96.05 (11.78). The intercept-only RVE model indicated a statistically significant overall effect size of 0.07,  $SE=0.04$ ,  $p=.000$ . Even though the overall effect size of 0.07 was statistically significant, the magnitude of the effect size was what Hopkins would describe as very small or trivial.

We examined the  $Q$ -value (an indicator of how similar the effects sizes are from study to study) and found a value of 1785.31, with  $df=139$  and  $p<.0001$ , indicating the true effect size varies from study to study. The  $I^2$  value reflects the proportion of variance that is due to real differences (and potentially explained by moderators). In this model,  $I^2$  is 92.214, which means that almost all the observed variance reflects real differences in study effects. However, when examining the risk of publication bias, we found there was good symmetry within the funnel plot, indicating no relationship between effect and study size; moreover, Egger's test indicated no evidence of bias ( $p=0.761$ ).

## Research Question 2

*What variables moderate the relationship between the TOSCRF and TOSWRF and other standardized measures of reading competence?*

In addition to an overall model, we examined whether study characteristics (i.e., domain of criterion test, type of learner, type of silent reading fluency test, and administration format) explained variability in effect sizes between the TOSWRF, TOSCRF, and other standardized measures of reading. The ROBUMETA package (Tanner-Smith & Tipton, 2014) was used to conduct tests based on RVE heteroscedasticity. Table 3 displays the results from meta-regression models predicting the average weighted effect sizes for all studies with four moderator variables.

When all four moderator variables were entered into the analysis, they accounted for 30% of the variance in effect sizes ( $Q$ -value = 75.42,  $df=10$ ,  $p=.000$ ). Three moderators (type of learner, ELL status, and type of silent reading fluency test) made a unique contribution to predicting variability in effect sizes. After controlling for other variables, the type of learner was a significant moderator variable. Specifically, English language learners made a small, unique, contribution to predicting variability in effect sizes between the TOSCRF, TOSWRF, and other standardized measures of reading,  $\beta = -.26$ ,  $SE = .09$ , 95% CI  $[-.43, -.09]$ ,  $p = .002$ , whereas students with and at risk for disabilities, average learners, and above average and gifted learners did not ( $ps > .05$ ).

We also found a small but significant difference in effect sizes between the TOSWRF and TOSCRF. The TOSCRF was responsible for larger effect sizes between standardized tests of reading competence than the TOSWRF,  $\beta = .23$ ,  $SE = .04$ , 95% CI  $[.16, .30]$ ,  $p = .000$ . Last, administration format significantly moderated effect sizes between the TOSWRF, TOSCRF and other standardized measures of reading competence. Tests administered to groups resulted in small, but statistically significant effects,  $\beta = -.11$ ,  $SE = .05$ , 95% CI  $[-.21, -.02]$ ,  $p = .016$ . There were no significant moderation effects for type of criterion tests, when effect sizes were compared between subtests (i.e., comprehension, fluency, vocabulary, word and letter identification, and decoding) and composite scores (overall reading scores or other composites related to achievement), after controlling for all other variables in the model.

## Discussion

In this meta-analysis, we sought to synthesize findings of two commonly used measures of silent reading fluency (i.e., the *Test of Silent Word Reading Fluency [TOSWRF]* and the *Test of Silent Contextual Reading Fluency [TOSCRF]*) and compare to other standardized measures of reading

competence to explore their usefulness as an indicator of general reading competence. A second objective was to examine variables that moderated variability in effect sizes on reading measures. Our findings contribute to the literature on using the TOSWRF and, its companion, the TOSCRF to measure reading competence.

### Research Question 1

*Do scores from the TOSWRF, the TOSCRF, and a weighted average of these two silent reading measures differ from those from other standardized measures of reading competence?*

The present meta-analysis included 47 published studies that administered the TOSWRF and/or the TOSCRF, and at least one of 48 separate standardized tests of reading, to 47,616 participants. The results from our investigation reveal several noteworthy findings. We found that scores from the TOSWRF, the TOSCRF, and a weighted average of these two silent reading measures differ very little from scores from other standardized measures of reading competence—the average standard score difference was 2 points.

The median average effect size difference between the TOSWRF and the subtest measures of component reading competence was .03 (very small or trivial) and ranged from  $-.16$  (small) to  $.23$  (small). The median average effect size difference between the TOSCRF and the tests of reading competence  $.16$  (small) and ranged from  $-.44$  (moderate) to  $.12$  (small). These results indicate that, on average, a student's performance on the TOSCRF and TOSWRF will mirror performance on other norm-referenced tests reading skills (and vice-versa), regardless of the type of component reading skill being measured.

The weighted mean differences between the silent reading fluency measures and composite scores from norm-referenced assessments (e.g., Gray Oral Reading Test [GORT], Kauffman Test of Educational Achievement [KTEA], and the Gates MacGinitie Reading Test [GMRT]) were larger than the subtests of component reading competence (Woodcock Johnson – III [WJ-III] Word Reading, Weschler Individual Achievement Test – III [WIAT-III] Word Reading, Peabody Picture Vocabulary Test [PPVT-4]). However, the effect size  $r$  values (i.e., the average differences) between the TOSCRF, TOSWRF, and standardized measures that produced composite scores and subtest scores were small ( $r = -.18$ , 95% CI  $[-.36, -.08]$ ) to trivial ( $r = -.04$ , 95% CI  $[-.16, .08]$ ), respectively (see Table 3).

The results also indicated that standardized mean differences ( $r$ ) between the TOSWRF, TOSCRF, and other norm-referenced, standardized tests of reading competence were trivial

**Table 3.** Meta-regression of Moderating Variables.

Moderator	$\beta$	SE	95% CI	$p$
Intercept	-.018	0.06	[-.13, .09]	.744
Type of Learner				
English Language Learner	-.261	0.09	[-.43, -.09]	.002
Average	.012	0.04	[-.06, .09]	.733
Above Average and/or Gifted	.392	0.21	[-.02, .81]	.064
Test of Silent Reading Fluency				
TOSWRF vs. TOSCRF	.228	0.04	[.16, .30]	.000
Administration Format				
Group vs. Individually Administered	-.114	0.05	[-.21, -.02]	.016
Type of Test Used for Criterion Validity				
Decoding	-.090	0.06	[-.21, .03]	.133
Word-Letter ID	.050	0.06	[-.07, .18]	.439
Fluency	.064	0.06	[-.06, .19]	.304
Vocabulary	-.011	0.12	[-.25, .23]	.927
Composite Score	.018	0.06	[-.10, .13]	.762

Note. All moderators were entered in one model. Several models were run for thorough subgroup comparisons for moderators with more than 2 categories. For the convenience of presentation, comparisons for all moderators are all listed in one table. Between-study sampling variance ( $\tau^2$ ) for this model is .03.

(weighted average  $r=0.07$ ); moreover, the range of effect size differences were relatively small (Shadish & Haddock, 2009; 0.04–0.11). What this suggests is that scores on the TOSWRF and TOSCRF are strong indicators of how well students are likely to perform across a broad range of reading competency tests.

## **Research Question 2**

*What variables moderate the relationship between the TOSCRF and TOSWRF and other standardized measures of reading competence?*

Our results showed that several moderators were responsible for differences in effect sizes across the 47 studies examined in this meta-analysis.

### **Domain of reading**

One finding was that type (or domain) of reading skill assessed by criterion reading measures did not explain a significant amount of variation in effect sizes. In other words, while true effect sizes varied across the 47 studies, the type of reading domain being assessed was not responsible for real differences in study effects. This confirmed that TOSCRF and/or TOSWRF scores do not differ significantly from scores across a wide range of domains (e.g., decoding, word identification, oral fluency, vocabulary, comprehension, and composite scores), which supports claims reported in the TOSCRF and TOSWRF technical manuals as well as existing theory suggesting that fluency scores are an indicator of general reading performance and can be useful as a progress monitoring measure (Espin & Deno, 2016; Hammill et al., 2014; Kamhi & Catts, 2017; Mather et al., 2014).

### **Type of learner**

Controlling for other variables, English language learner was associated with a small, negative moderating effect ( $\beta = -.26$ ,  $SE = .09$ ) on the standardized mean differences ( $r$ ) between the TOSWRF, TOSCRF, and other standardized measures of reading. The range in effect size differences was  $-.43$  (moderate) to  $-.09$  (very small to trivial). Further analysis revealed English language learners earned higher average mean scores on the TOSWRF and TOSCRF ( $M = 85.25$ ,  $SD = 21.90$ ) than on the other standardized tests of reading ( $M = 78.53$ ,  $SD = 22.75$ ); a mean difference of 6.72 points. This finding is consistent with what has been reported in prior literature, showing that standardized assessment results for ELLs are confounded by English language proficiency, with the largest performance differences in language-related subscales of tests of reading (Abedi, 2002).

Other studies (e.g., Denton et al., 2011; Hua & Keenan, 2017) suggest that the factors affecting students' performance on different reading assessments varies by the reading skill level of the student. The current analysis did not find moderation by skill level suggesting that student performance on the TOSCRF and TOSWRF varies in concert with other reading measures for differing ability levels. This does not contradict others' findings that assessment varies in its effectiveness based on the ability of the student being tested. Rather, this finding indicates that the TOSWRF and TOSCRF are like other measures in their level of utility for students of varying abilities. Future study will be needed to determine if the TOSWRF and TOSCRF can detect changes in students' performance in response to instruction, especially for lower-performing students.

### **Administration format**

The type of silent reading assessment (TOSCRF v. TOSWRF) also had a small, moderating effect, on mean differences between tests. While the magnitude of differences were small, weighted

average scores between the TOSCRF and other standardized reading tests tended to be slightly larger than differences on the TOSWRF ( $M = 5.27$ ,  $SD = 0.09$ ). These findings seem to align with research showing that mastery of word reading precedes students' reading of connected text (Altani et al., 2020; Jenkins et al., 2003), which would translate into higher overall scores on tests that measure word reading fluency. In addition, the administration format moderator showed that effect sizes across studies varied based on whether standardized reading tests were administered to individuals or groups. This finding was consistent with what has been reported in other meta-analytic studies examining curriculum-based measurement oral reading as an indicator of reading achievement (Reschly et al., 2009).

### **Implications**

These findings have implications for school professionals. To elaborate, one should consider that both tests can be administered by a teacher to a classroom of students in two 3-minute sessions. This feature makes the TOSWRF and TOSCRF potentially attractive measures of student progress after initial diagnostic assessment and targeted instruction. Importantly, we are not suggesting that the TOSWRF and TOSCRF should replace traditional standardized reading measures. However, the two might be used together as a supplementary tool within screening systems such as the Response to Intervention (RTI) model, specifically for tracking intervention effectiveness more efficiently.

For teachers and reading specialists, the implication seems to be that language clearly plays a central role in how nonnative speakers of English perform on the TOSWRF and TOSCRF; moreover, since many ELLs are just learning English, the best use of classroom time might be to help ELLs develop English word recognition and decoding skills and assess read-aloud fluency using measures such as DIBELS Oral Reading Fluency (ORF) where the learner can hear and decode written language aloud. As nonnative speakers' reading abilities evolve, teachers might consider integrating tests of silent reading fluency into formal and informal screening and progress monitoring procedures.

### **Limitations**

The study has several limitations that should be mentioned. First, although we searched numerous databases to collect studies for inclusion, we might have missed some applicable studies. Some of the studies in the review also had more effect sizes than other studies. To mitigate the potential problems associated with dependencies in nested data, we used RVE analysis procedures (Hedges et al., 2010). Second, nearly 60% of the studies in the meta-analytic sample were performed with students who were average-achieving students. Less than 5% of the sample included either English language learners or students who were academically gifted. Therefore, more work is needed to better understand the efficacy of tests of silent reading fluency and other standardized measures with non-native English speakers as well as with students who are academically gifted.

An important objective of this investigation was to identify critical variables in studies (i.e., domain used for criterion validity, type of learner, administration format) that may explain excess variability in effect sizes between the TOSWRF, TOSCRF and other standardized measures of reading. However, due to inconsistencies in reporting across studies, we were unable to consistently collect data on age and/or grade-level, which can impact performance on silent reading measures (Kim et al., 2011; Price et al., 2016). Systematic reviews on oral reading fluency report inconsistent data collection for age and/or grade-level as well (Reschly et al., 2009); therefore, more work is needed to better understand how participants' age or grade-level affects performance across studies. Last, some may disagree about how we classified each of the criterion reading



measures in the study. Admittedly, the constructs that most reading tests measure are not without overlap. For example, the TOWRE Word Reading Efficiency score is a combination score of a decoding task and a word recognition task. Both tasks require fluent word production and therefore could fit in four different categories (decoding, word identification, fluency, or composite score).

## Conclusion

The TOSWRF and its companion test, the TOSCRF, were built to provide educational professionals (e.g., researchers, school psychologists, and teachers) with time- and cost-efficient, reliable, and valid indicators of reading competence. In this meta-analysis, the results from administering the TOSWRF and TOSCRF are nearly identical to the results from administering other standardized measures of reading competence. These findings seem to hold across types of reading competence and with most types of students. Furthermore, the results suggest that these brief, reliable measures of silent reading fluency perform at least as well as more traditional, resource-intensive measures when screening to identify students who have reading problems or when a brief, accurate, easy to administer and score measure is required.

## Disclosure statement

No potential conflict of interest was reported by the author(s).

## ORCID

Daniel R. Wissinger  <http://orcid.org/0000-0003-3615-096X>  
 Adrea J. Truckenmiller  <http://orcid.org/0000-0002-6101-6175>

## References

- Abedi, J. (2002). Standardized achievement tests and English language learners: Psychometrics issues. *Educational Assessment*, 8(3), 231–257. [https://doi.org/10.1207/s15326977ea0803\\_02](https://doi.org/10.1207/s15326977ea0803_02)
- Allinder, R. M., & Eccarius, M. A. (1999). Exploring the technical adequacy of curriculum-based measurement in reading for children who use manually coded English. *Exceptional Children*, 65(2), 271–283. <https://doi.org/10.1177/001440299906500210>
- Altani, A., Protopapas, A., Katopodi, K., & Georgiou, G. K. (2020). From individual word recognition to word list and text reading fluency. *Journal of Educational Psychology*, 112(1), 22–39. <https://doi.org/10.1037/edu0000359>
- Armbruster, B. B. (2010). *Put reading first: The research building blocks for teaching children to read, kindergarten through Grade 3*. Center for the Improvement of Early Reading Achievement (CIERA).
- Berendes, K., Wagner, W., Meurers, D., & Trautwein, U. (2019). When silent reading fluency test measures more than reading fluency: Academic language features predict the test performance of students with non-German home language. *Reading and Writing*, 32(3), 561–583. <https://doi.org/10.1007/s11145-018-9878-x>
- Borenstein, M., Hedges, L. V., Higgins, J. P., & Rothstein, H. R. (2009). *Introduction to meta-analysis*. Wiley.
- Cirino, P. T., Romain, M. A., Barth, A. E., Tolar, T. D., Fletcher, J. M., & Vaughn, S. (2013). Reading skill components and impairments in middle school struggling readers. *Reading and Writing*, 26(7), 1059–1086. <https://doi.org/10.1007/s11145-012-9406-3>
- Cohen, J. (1988). *Statistical power for the behavioral sciences* (2nd ed.). Erlbaum.
- Connor, C. M., Alberto, P. A., Compton, D. L., & O'Connor, R. E. (2014). *Improving reading outcomes for students with or at risk for reading disabilities: A synthesis of the contributions from the Institute of Education Sciences Research Centers*. National Center for Special Education Research, Institute of Education Sciences, U.S. Department of Education. <http://ies.ed.gov/>
- Denton, C. A., Barth, A. E., Fletcher, J. M., Wexler, J., Vaughn, S., Cirino, P. T., Romain, M., & Francis, D. J. (2011). The relations among oral and silent reading fluency and comprehension in middle school: Implications for identification and instruction of students with reading difficulties. *Scientific studies of Reading*, 15(2), 109–135. <https://doi.org/10.1080/10888431003623546>

- Egger, M., Smith G., Schneider, M., & Minder, C. (1997). Bias in meta-analysis detected by a simple, graphical test. *British Medical Journal*, 315(7109), 629–634.
- Espin, C. A., & Deno, S. L. (2016). Oral reading fluency or reading aloud from text: An analysis through a unified view of construct validity. In K. Cummings & Y. Petscher (Eds.), *The fluency construct* (pp. 365–383). Springer.
- Finger, M. S., & Ones, D. S. (1999). Psychometric equivalence of the computer and booklet forms of the MMPI: A meta-analysis. *Psychological Assessment*, 11(1), 58–66. <https://doi.org/10.1037/1040-3590.11.1.58>
- Freeland, J., Skinner, C., Jackson, B., McDaniel, C. E., & Smith, S. D. (2000). Measuring and increasing silent comprehension rates: Empirically validating a repeated readings intervention. *Psychology in the Schools*, 37(5), 415–429. [https://doi.org/10.1002/1520-6807\(200009\)37:5<415::AID-PITS2>3.0.CO;2-L](https://doi.org/10.1002/1520-6807(200009)37:5<415::AID-PITS2>3.0.CO;2-L)
- Fuchs, L. S., Fuchs, D., Hosp, M. K., & Jenkins, J. R. (2001). Oral reading fluency as an indicator of reading competence: A theoretical, empirical, and historical analysis. *Scientific Studies of Reading*, 5(3), 239–256. <https://doi.org/10.4324/9781410608246-3>
- Fuchs, L. S. (2004). The past, present, and future of curriculumbased measurement research. *School Psychology Review*, 33(2), 188–192.
- Hammill, D. D., Wiederholt, J. L., & Allen, E. A. (2006). TOSCRF: Test of silent contextual reading fluency: Examiner's manual. Austin, TX: PRO-ED.
- Hammill, D. D., Wiederholt, J. L., & Allen, E. A. (2014). *Test of silent contextual reading fluency*. (2nd ed.). PRO-ED.
- Hedges, L. V., & Olkin, I. (1985). *Statistical methods for meta-analysis*. Academic Press.
- Hedges, L.V., Tipton, E., & Johnson, M.C. (2010). Robust variance estimation in meta-regression with dependent effect size estimates. *Research Synthesis Methods*, 1(1), 39–65.
- Hedges, L. V., Tipton, E., & Johnson, M. C. (2010). Erratum: Robust variance estimation in meta-regression with dependent effect size estimates. *Research synthesis Methods*, 1(2), 164–165. <https://doi.org/10.1002/jrsm.17>
- Hiebert, E. (2015). *Teaching stamina and silent reading in the digital-globe age*. TextProject.
- Hiebert, E. H., Samuels, S. J., & Rasinski, T. (2012). Comprehension-based silent reading rates: What do we know? What do we need to know? *Literacy Research and Instruction*, 51(2), 110–124. <https://doi.org/10.1080/19388071.2010.531887>
- Higgins, J., & Green, S. (2011). *Cochrane handbook for systematic reviews of interventions*. Wiley-Blackwell.
- Hopkins, W. G. (2002). Probabilities of clinical or practical significance', *Sportscience*, 6, p. Available at: [sportsci.org/jour/0201/wghprob.htm](http://sportsci.org/jour/0201/wghprob.htm)
- Hua, A. N., & Keenan, J. M. (2017). Interpreting reading comprehension test results: Quantile regression shows that explanatory factors can vary with performance level. *Scientific Studies of Reading*, 12(3), 225–238. <https://doi.org/10.1080/10888438.2017.1280675>
- Jenkins, J. R., Fuchs, L. S., Van Den Broek, P., Espin, C., & Deno, S. L. (2003). Sources of individual differences in reading comprehension and reading fluency. *Journal of Educational Psychology*, 95(4), 719–729. <https://doi.org/10.1037/0022-0663.95.4.719>
- Kamhi, A. G., & Catts, H. W. (2017). Epilogue: Reading comprehension is not a single ability – Implications for assessment and instruction. *Language, Speech, and Hearing Services in Schools*, 48(2), 104–107. [https://doi.org/10.1044/2017\\_LSHSS-16-0049](https://doi.org/10.1044/2017_LSHSS-16-0049)
- Kara, Y., Kamata, A., Potgieter, C., & Nese, J. F. (2020). Estimating model-based oral reading fluency: A Bayesian approach. *Educational and Psychological Measurement*, 80(5), 847–869.
- Kim, Y.-S., Wagner, R. K., & Foster, E. (2011). Relations among oral reading fluency, silent reading fluency, and reading comprehension: A latent variable study of first-grade readers. *Scientific studies of Reading*, 15(4), 338–362. <https://doi.org/10.1080/10888438.2010.493964>
- Lembke, E., Hampton, D., & Hendrick, E. (2013). Data-based decision-making in academics using curriculum-based measurement. In J. W. Lloyd, T. J. Landrum, B. Cook, & M. Tankersley (Eds.), *Research-based approaches for assessment* (pp. 18–31). Pearson.
- Mather, N., Hammill, D. D., Allen, E. A., & Roberts, R. (2004). *Test of silent word reading fluency*. PRO-ED.
- Mather, N., Hammill, D. D., Allen, E. A., & Roberts, R. (2014). *Test of silent contextual reading fluency* (2nd ed.). PRO-ED.
- Mills, G. E., & Gay, L. R. (2019). *Educational research: Competencies for analysis and applications* (12th ed.). Pearson.
- Price, K. W., Meisinger, E. B., Louwarse, M. M., & D'Mello, S. K. (2012). Silent reading fluency using underlining: Evidence for an alternative method of assessment. *Psychology in the Schools*, 49(6), 606–618. <https://doi.org/10.1002/pits.21613>
- Price, K. W., Meisinger, E. B., Louwarse, M. M., & D'Mello, S. (2016). The contributions of oral and silent reading fluency to reading comprehension. *Reading Psychology*, 37(2), 167–201. <https://doi.org/10.1080/02702711.2015.1025118>

- Rasinski, T., Rupley, W., Paige, D., & Nichols, W. (2016). Alternative text types to improve reading fluency for competent to struggling readers. *International Journal of Instruction*, 9(1), 163–178. <https://doi.org/10.12973/iji.2016.9113a>
- Rasinski, T., Samuels, S. J., Hiebert, E., Petscher, Y., & Feller, K. (2011). The relationship between a silent reading fluency instructional protocol on students' reading comprehension and achievement in an urban school setting. *Reading Psychology*, 32(1), 75–97. <https://doi.org/10.1080/02702710903346873>
- Reschly, A. L., Busch, T. W., Betts, J., Deno, S. L., & Long, J. D. (2009). Curriculum-based measurement oral reading as an indicator of reading achievement: A meta-analysis of the correlational evidence. *Journal of School Psychology*, 47(6), 427–469. <https://doi.org/10.1016/j.jsp.2009.07.001>
- Reutzel, D. R., & Juth, S. (2014). Supporting the development of silent reading fluency: An evidence-based framework for the intermediate grades (3–6). *International Electronic Journal of Elementary Education*, 7(1), 27–46. <https://files.eric.ed.gov/fulltext/EJ1053594.pdf>
- Ritchey, K. D., McMaster, K. L., Otaiba, S. A., Puranik, C. S., Grace Kim, Y. S., Parker, D. C., & Ortiz, M. (2016). Indicators of fluent writing in beginning writers. In K. Cummings & Y. Petscher (Eds.), *The fluency construct*. Springer.
- Samuels, S. J., Hiebert, E., & Rasinski, T. (2015). Eye movements make reading possible. In E. Hiebert (Ed.), *Teaching Stamina and silent reading in the digital-globe age* (pp. 32–57). TextProject.
- Shadish, W. R., & Haddock, C. K. (2009). Combining estimates of effect size. In H. Cooper, L. V. Hedges, & J. C. Valentine (Eds.), *The handbook of research synthesis and meta-analysis* (pp. 257–277). Russell Sage Foundation.
- Shin, J., & McMaster, K. (2019). Relations between CBM (oral reading and maze) and reading comprehension on state achievement tests: A meta-analysis. *Journal of School Psychology*, 73(1), 131–149. <https://doi.org/10.1016/j.jsp.2019.03.005>
- Tanner-Smith, E. E., & Tipton, E. (2014). Robust variance estimation with dependent effect sizes: Practical considerations including a software tutorial in Stata and SPSS. *Research synthesis Methods*, 5(1), 13–30. <https://doi.org/10.1002/jrsm.1091>
- Turney, J. (1977). *Exploratory data analysis*. Addison Wesley.
- Van den Boer, M., Bazen, L., & de Bree, E. (2022). The same yet different: Oral and silent reading in children and adolescents with dyslexia. *Journal of Psycholinguistic Research*, 51, 803–817. <https://doi.org/10.1007/s10936-022-09856-w>
- Wagner, R. K., Torgesen, J. K., Rashotte, C. A., & Pearson, N. A. (2010). *Test of silent reading efficiency and comprehension*. PRO-ED.