A Meta-Analysis on the Relation between Reading and Working Memory

Abstract

Background.

Working memory (WM) refers to the capacity to store information for short periods of time while engaging in cognitively demanding activities (Baddeley, 1986). Theoretically, WM plays an important role in reading performance because many reading tasks involve simultaneous information processing and storage. For example, to comprehend a passage, individuals first visually process the words; then match the words to the phonological, orthographic, and semantic representations in long-term memory; and finally combine these representations with the context to create an understanding of the passage. WM is purported to be involved in this whole reading process by keeping relevant information in short-term memory, retrieving information from long-term memory, and integrating all information to form an accurate representation of the situation described by the text (Christopher et al., 2012; van den Broek, Mouw, & Kraal, 2016). However, despite the theoretically strong relation between reading and WM, substantial differences have been reported in the percentage of variance in reading that is explained by WM. Some studies have found that the contribution of WM to reading is negligible, with $R^2$ around 0 (e.g., Koltun, 2003; O'Shaughnessy & Swanson-1997), while other studies have found that WM almost fully explained reading performance, with $R^2$ values ranging from .60 to .81 (e.g., Daneman, 1980; Weissinger, 2013; McIntyre, 2015).

Clearly, it is important to gain a better insight into the degree to which WM accounts for reading performance and the factors that influence this relation. To our knowledge, two prior reviews aimed to investigate these issues (i.e., Daneman & Merikle, 1996; Savage, Lavers, & Pillay, 2007). Specifically, Savage et al. (2007) did a comprehensive literature review on the relation between reading and WM. Their major findings were (a) in comparison to foundational reading skills such as decoding, WM is more important for advanced reading skills such as reading comprehension even when other reading skills such as decoding and vocabulary are controlled for; and (b) whether the relation between reading and WM is influenced by domains of WM (i.e., verbal vs. visual-spatial) remains unclear. However, the Savage et al. (2007) study is a comprehensive literature review, which warrants updating and systematic meta-analytic investigations of the studies cited in that literature review as well as studies published since this review. In contrast to Savage et al. (2007), Daneman and Merikle (1996) used the meta-analytic method to systematically investigate the relation between WM and language comprehension. Daneman and Merikle (1996) found that WM, measured using dual-task paradigms (e.g., listening span; backward digit recall), is a better predictor of comprehension than short-term memory. In their meta-analysis, performance on WM tasks tapping both verbal and numerical domains, in contrast to WM tasks tapping other domains (e.g., verbal or visuospatial), best predicted comprehension as indexed by vocabulary, listening comprehension, and reading comprehension. Although Daneman and Merikle (1996) adopted the meta-analytic method, they only focused on language comprehension and did not investigate the relation between WM and different comprehension skills (i.e., vocabulary, listening comprehension, reading comprehension) or the relation between WM and foundational reading skills such as phonological coding and decoding. Moreover, neither Savage et al. (2007) nor Daneman and Merikle (1996) sufficiently addressed the issue that whether WM makes unique contributions to reading comprehension when relevant foundational reading skills are controlled for, nor were they able to determine whether the relations of WM and reading change across development.

Purpose.

The present study aims to address the issues mentioned above. We use the meta-analytic method to systematically investigate the relation between reading and WM with a focus on several moderators that potentially explain variations in this relation. These moderators include domains of WM (i.e., verbal WM, numerical WM, visuospatial WM), types of reading (i.e., phonological coding, decoding, vocabulary, comprehension), and grade level (i.e., before 4th grade vs. at/beyond 4th grade). Moreover,
we examined the unique contribution of WM to reading comprehension after controlling for decoding and vocabulary.

**Research Design.** Articles for this meta-analysis were identified in two ways. First, a computer search of the ProQuest, PsycARTICLES, PsycINFO databases for literature was conducted. We used the earliest possible start date (1964) through October 2015. Titles, abstracts, and keywords were searched for the following terms: *working memory* AND *read*\(^*\), *word*, *decod*\(^*\), *phonolog*\(^*\), *comprehend*\(^*\), *vocabulary*, OR *language*. The terms *read*\(^*\), *decod*\(^*\), *phonolog*\(^*\), and *comprehend*\(^*\) allowed for inclusion of reading, phonology/phonological, comprehend/compression, and so forth. Second, we hand searched citations in prior relevant reviews (Daneman & Merikle, 1996; Savage, Lavers & Pillay, 2007). We searched unpublished literature through Dissertation and Masters Abstract indexes in ProQuest, Cochrane Database of Systematic Reviews, relevant conference programs (e.g., Society for the Scientific Study of Reading, Conference of Society for Research on Educational Effectiveness, and Annual Conference of American Educational Research Association), and emailing some researchers likely to have conducted work in this area. We also contacted several researchers to request for correlation tables not provided in studies. The initial search yielded 69,007 studies. The second author and the fifth author then reviewed all studies by titles and abstract. After excluding the duplicate 86 articles, and 66,560 irrelevant articles, the remaining 2361 articles were closely reviewed using the specific criteria described below (also see Figure 1 for the flow diagram for the search and inclusion criteria for studies in this review).

![Flow diagram](image)

**Figure 1.** Flow diagram for the search and inclusion criteria for studies in this review.

First, studies had to include at least one quantitative task measuring WM and at least one quantitative task measuring reading. WM measures refer to the tasks that tap processing and storage simultaneously (e.g., complex span tasks and dual-tasks WM). Measures that tap executive functions,
such as inhibition, switching, or updating, were not considered as WM measures in this meta-analysis. Reading measures refer to the tasks that tap one of the following skills: phonological coding, decoding (word recognition and non-word reading), fluency, vocabulary, listening comprehension, reading comprehension, and comprehensive reading that tap at least two of the above-mentioned reading skills. To be considered as a measure of phonological coding, the task must involve deletion, blending, counting, segmentation, generation, judgment, position analysis or replacement of phoneme, onset, rhymes, and/or syllables in words. To be considered as a fluency measure, the test needed to measure timed word, non-word, sentence, passage reading accuracy. To be considered as a decoding measure, the test had to comprise reading accuracy of words, non-words, sentences, or passages. To be considered a measure of reading comprehension, studies in which a child read a passage or sentence and answered questions in relation to the text were included. To be considered a measure of listening comprehension, tests that measured understanding of heard sentences or passages by means of oral cloze, answering questions in relation to the orally presented text, or retelling the orally presented text were included. To be considered a measure of vocabulary, tests should measure expressive or receptive vocabulary using pictures and measures of synonyms or antonyms. Table 1 demonstrates the examples of WM measures and different reading measures considered in this study.

Second, studies had to report at least one correlation ($r$) between any measure of WM and any measure of reading, or the percentage of variance ($R^2$) in reading accounted for by WM only. The measures of WM and reading used to calculate the direct correlation (not partial correlation) had to be taken at the same time point, because we were interested in the concurrent direct relation between reading and WM and how this relation was affected by the moderators proposed in this study.

Third, we only included studies that focused on reading in English (provided measures on verbal WM in English and reading in English). We only focused on general populations (an unselected sample), because the heterogeneity of disability groups is often not systematically studied or reported in many primary studies. Also, because aging influences human cognition including WM, we did not include studies or findings from studies that are based on adults over 50 years of age.

Coding Procedure and Inter-rater Reliability

Studies were coded according to the characteristics of participants and tasks used to measure WM and reading. Not all studies provided sufficient information on the variables of interest for the present study. In case of insufficient information, authors were contacted to obtain the missing information. In addition to these variables, we also coded the number of participants ($N$) used to obtain each correlation. The latter was needed to weight each effect size, so that correlations obtained from larger samples were given more weight in the analysis than those obtained from smaller samples. The important features of individual studies are provided in Appendix A.

Variables were discussed until a consensus was reached between the first and the second authors. Then, both the first author and the second author used this coding system to conduct the final coding of all 198 studies independently. The inter-rater reliability was 1 for publication type, .98 for grade level, .97 for sample size, .95 for correlation coefficients, .97 for domains of WM, .95 for types of reading skills. Any disagreements between raters were resolved by consulting the original article or by discussion.

Missing data. During coding, many instances of missing data became apparent. If the data that were missing were critical for calculating the effect size for the main outcomes, we contacted the author. However, this approach was usually unsuccessful. If data were missing for moderator variables, the study was excluded from the moderator analysis for which data were missing but was included in all moderator analyses for which data were provided.

Analytic Strategies

The effect size index used for all outcome measures was Pearson’s $r$, the correlation between reading and WM. We considered all eligible effect sizes in each study. That is, studies could contribute multiple effect sizes as long as the sample for each effect size was independent. For studies that reported
multiple effect sizes from the same sample, we accounted for the statistical dependencies using the random effects robust standard error estimation technique developed by Hedges, Tipton, and Johnson (2010). This analysis allowed for the clustered data (i.e., effect sizes nested within samples) by correcting the study standard errors to take into account the correlations between effect sizes from the same sample. The robust standard error technique requires that an estimate of the mean correlation ($\rho$) between all the pairs of effect sizes within a cluster be estimated for calculating the between-study sampling variance estimate, $\tau^2$. In all analyses, we estimated $\tau^2$ with $\rho = .80$; sensitivity analyses showed that the findings were robust across different reasonable estimates of $\rho$.

Analyses were based on Borenstein, Hedges, Higgins, and Rothstein’s (2005) recommendations. Specifically, we converted the correlation coefficients to Fisher’s Z scale, and all analyses were performed using the transformed values. The results, such as the summary effect and its confidence interval, were then converted back to correlation coefficients for presentation. Also, because we hypothesized that this body of research reports a distribution of correlation coefficients with significant between-studies variance, as opposed to a group of studies that attempts to estimate one true correlation, a random-effects model was appropriate for the current study (Lipsey & Wilson, 2001). Weighted, random-effects meta-regression models using Hedges et al.’s (2010) corrections were run with ROBUMETA in Stata (Hedberg, 2011) to summarize correlation coefficients and to examine potential moderators. Specifically, we first estimated only the overall weighted mean correlation between WM and reading.

Then, subgroup analyses were used to examine the relation between reading and WM for each subgroup of each moderator. Meta-regression analyses were used to examine whether domains of WM, types of reading skills, and grade level moderated the relation between reading and WM. For the moderation analysis, all moderators were entered into the model simultaneously, with publication type (peer reviewed articles vs. other types of publications) as the covariate in the model as well. Because the moderators were all categorical, we created several sets of dummy coded variables to examine the comparisons among categories (Cohen, Cohen, West, & Aiken, 2013).

For the analysis on the unique contributions of WM to reading comprehension, we calculated the partial correlations based on correlation matrices of studies that provided the correlations among 1) WM, decoding, and reading comprehension; 2) WM, vocabulary, and reading comprehension; or 3) WM, decoding, vocabulary, and reading comprehension. For each study, the correlation matrix was recorded along with the means, standard deviations, and number of observations. For example, if a study reported 2 measures on decoding, 2 measures on reading comprehension, and 2 measures on WM, then the full 2 X 2 X 2 correlation matrix was recorded, producing 6 partial correlations between WM and decoding (partialling out decoding).

We then synthesized these partial correlations to indicate the unique contribution of WM to reading comprehension, partialling out decoding or vocabulary or both. Because the way we calculated partial correlations produces many effect sizes nested within a sample, we accounted for the statistical dependencies using the random effects robust standard error estimation technique developed by Hedges, Tipton, and Johnson (2010) as mentioned earlier. In all analyses, we estimated $\tau^2$ with $\rho = .80$; sensitivity analyses showed that the findings were robust across different reasonable estimates of $\rho$.

Publication bias (the problem of selective publication, in which the decision to publish a study is influenced by its results) was examined using the method of Egger, Smith, Schneider, and Minder (1997) and funnel plot. We did not find significant publication bias based on Egger et al.’s (1997) publication bias statistics (i.e., the standard errors of correlations did not significantly predict correlations among studies with ROBUMETA in Stata, $ps > .11$), except for the correlation between phonological coding and WM, $p = .01$. Further funnel plot analyses showed reasonable symmetry in all reported correlations (the significant Egger’s test for the correlation between phonological coding and WM may be due to two outliers in the funnel plot). Taken together, Egger’s test and funnel plot suggest that there was little influence of publication bias in the data and thus the original dataset was used in all reported analyses.
Findings.

We found that the relation between reading and WM was medium \((r = .29)\) and was significantly influenced by domains of WM and grade level. Specifically, before 4th grade, different domains of WM related to reading to a similar degree, whereas verbal WM showed stronger relation with reading than numerical WM and visuospatial WM at/beyond 4th grade. WM showed a stronger relation to reading, especially with comprehension and vocabulary, before 4th grade than that at/beyond 4th grade. Types of reading skills did not appear to influence the relation between reading and WM; with the one exception for decoding, where we found that word recognition showed stronger relations with WM than that of non-word reading and word list reading accuracy showed stronger relations with WM than that of sentence/passage reading accuracy. Our findings indicate that WM makes unique contributions to reading comprehension beyond decoding or vocabulary; however, WM makes no unique contributions to reading comprehension when we control for both decoding and vocabulary.