


A Meta-Analysis of the Long-Term Effects of Phonemic Awareness, Phonics, Fluency, and Reading Comprehension Interventions

Journal of Learning Disabilities
2016, Vol. 49(1) 77–96
© Hammill Institute on Disabilities 2014
Reprints and permissions:
sagepub.com/journalsPermissions.nav
DOI: 10.1177/0022219414528540
journaloflearningdisabilities.sagepub.com


Sebastian P. Suggate, PhD¹

Abstract

Much is known about short-term—but very little about the long-term—effects of reading interventions. To rectify this, a detailed analysis of follow-up effects as a function of intervention, sample, and methodological variables was conducted. A total of 71 intervention-control groups were selected ($N = 8,161$ at posttest) from studies reporting posttest and follow-up data ($M = 11.17$ months) for previously established reading interventions. The posttest effect sizes indicated effects ($d_w = 0.37$) that decreased to follow-up ($d_w = 0.22$). Overall, comprehension and phonemic awareness interventions showed good maintenance of effect that transferred to nontargeted skills, whereas phonics and fluency interventions, and those for preschool and kindergarten children, tended not to. Several methodological features also related to effect sizes at follow-up, namely experimental design and dosage, and sample attrition, risk status, and gender balance.

Keywords

reading intervention, long-term, comprehension, decoding, phonics, phonemic awareness, fluency

Previous studies and meta-analyses have shown that reading interventions can be an effective means to improve children's reading skills in the short-term (e.g., Bus & van Ijzendoorn, 1999; Ehri, Nunes, Stahl, & Willows, 2001; Ehri, Nunes, Willows, et al., 2001; Suggate, 2010; Swanson, Hoskyn, & Lee, 1999). However, it is clearly of great interest to both researchers and practitioners to move beyond the study of short-term effects to understand not only whether reading interventions result in longer-term gains (Blachman et al., 2014), but also which features of reading interventions relate to intervention outcomes.

Reading Interventions and Reading Problems

The etiology of reading problems is diverse and controversial. Generally, it is accepted that children can experience difficulty with some combination of (a) semantic and (b) phonological aspects of reading (Coltheart, Curtis, Atkins, & Haller, 1993; Gough & Tunmer, 1986; Hatcher, Hulme, & Ellis, 1994; Nation & Coxsey, 2009). Difficulty and diversity in categorizing reading problems occur through disagreement about whether discrepancy definitions are appropriate or whether low reading achievement alone suffices (Bell, McCallum, & Cox, 2003; Ferrer, Shaywitz, Holahan, Marchione, & Shaywitz, 2010; Harm & Seidenberg, 1999; Tunmer & Greaney, 2009). Still further distinctions arise from a response to intervention perspective (e.g., Scholin & Burns, 2012). Here, no attempt is made

to weigh in on this debate, but to gather existing categories from research and subject these to meta-analytical investigation. Based on reading research conducted in the past decades (e.g., see studies cited in the appendix) and placing precocious readers to one side (Stainthorp & Hughes, 2004), the following samples are frequently mentioned in reading intervention literature: (a) normal readers; (b) at-risk readers, usually either reading below the 50th percentile or originating from socially or economically disadvantaged groups including second language learners; (c) low-performing readers, usually reading below the 25th percentile; and (d) reading disabled students, either reading below the 10th percentile or those who have been diagnosed as having a reading-IQ discrepancy of one standard deviation. Given the centrality of reading to all citizens in society, a fifth category might be added, whereby children have a learning or cognitive disability.

Previous work has begun to investigate the effect of sample risk status on reading intervention effectiveness in the short term (e.g., Ehri, Nunes, Willows, et al., 2001; Suggate, 2010; Swanson et al., 1999), but has provided mixed results

¹University of Regensburg, Germany

Corresponding Author:

Sebastian P. Suggate, Department of Education, Alanus University, Villestrasse 3, 53347 Alfter near Bonn, Germany.
Email: Sebastian.Suggate@alanus.edu

and, with regard to long-term effects, virtually no results. For example, Ehri, Nunes, Stahl, et al. (2001) found that phonics helped younger at-risk readers more than younger readers with a more severe impairment, whereas phonemic awareness helped all readers equally (Ehri, Nunes, Willows, et al., 2001). In terms of comprehension interventions, these have been shown to be effective for older disabled students, in particular (Talbot, Wills, & Tankersley, 1994; Wanzek, Vaughn, Kim, & Cavanaugh, 2006). Considering both comprehension and phonetic-decoding interventions together, Suggate (2010) did not find that at-risk or more struggling readers differentially benefitted; however, he included only disadvantaged readers, whereas including normally achieving readers provides an important reference point in deciding how effective interventions for disadvantaged readers are. Of interest, Scholin and Burns (2012) found no links between preintervention level and postintervention growth on various reading tests, suggesting that risk status determined at pretest might not relate to responsiveness to intervention. In summary, evidence indicates that reading interventions generally benefit all readers, although research is needed investigating effects at a more long-term follow-up to test whether and how different readers respond to reading intervention.

Intervention Type

Previous meta-analyses suggest that, in particular, phonemic awareness and phonics interventions are particularly helpful for younger children (Bus & van Ijzendoorn, 1999; Ehri, Nunes, Stahl, et al., 2001; Ehri, Nunes, Willows, et al., 2001). Consistent with Ehri, Nunes, Stahl, et al. (2001), phonological awareness interventions are defined as those that increase children's awareness of the sounds at the word level (e.g., dig, dug, dog). Phonemic awareness interventions target awareness of the sounds (i.e., phonemes) composing words (e.g., "cat" as /k/ /a/ /t/). Accordingly, phonemic awareness is more specific to reading because this often requires decoding words at the phoneme level. Phonics interventions teach associations between phonemes and orthography, thereby they differ from pure phonemic awareness interventions in that they directly incorporate letters or text. Fluency interventions target "the ability to read with speed and fluency" (Therrien, 2004, p. 252) and generally include repeated reading, tutoring, or peer-reading activities (Fuchs & Fuchs, 2005).

Turning to interventions with a lesser focus on phoneme and text level decoding, reading comprehension interventions provide "specific procedures that guide students to become aware of how well they are comprehending as they attempt to read" (National Reading Panel, 2000, pp. 4–39). Typical activities in reading comprehension interventions, involve reflection, prior knowledge, question generation, pictorial cues, identifying themes, inferential thinking, summarization, and story structure (Suggate, 2010). Such

comprehension interventions have also been shown to relate positively to intervention outcomes (Elbaum, Vaughn, Hughes, & Moody, 2000; National Early Literacy Panel, 2008; National Reading Panel, 2000; Suggate, 2010; Swanson et al., 1999; Talbot et al., 1994). Typically, comprehension interventions are provided to older students who can already decode; however, one notable exception is Reading Recovery (Center, Wheldall, Freeman, Outhred, & McNaught, 1995) and similar interventions based on a whole language approach (Suggate, 2010). These interventions focus on teaching strategies to infer both word and sentence meaning and also to decode words based on surrounding contextual information. A further feature of these early comprehension interventions is that sound-to-spelling instruction is either absent altogether or conducted in an incidental manner (Buckingham, Wheldall, & Beaman, 2012; Tunmer, Chapman, & Prochnow, 2004).

Of interest, the phonological linkage hypothesis (Hatcher, Hulme, & Ellis, 1994) predicts that phonics interventions would show an advantage of phonemic awareness interventions, by virtue of their providing explicit links between phonemes and words—a hypothesis that has not yet been born out by short-term meta-analytical reviews (see Ehri, Nunes, Stahl, et al., 2001; Ehri, Nunes, Willows, et al., 2001). However, it is conceivable that skills instructed in one intervention type or the other might show differential transfer effects and maintenance to follow-up, if for example they are less able to be integrated into the reading process, or if they represent skills that children acquire in the interim without the intervention (e.g., Paris, 2005).

Suggate (2010) grouped the above reading interventions into two categories, namely (a) phonetic decoding, which contained phonological, phonemic, phonics, and fluency interventions, and (b) comprehension interventions, including whole-language approaches. Of interest, it was found that phonetic-decoding interventions were more effective for children in kindergarten and Grade 1, with comprehension interventions more effective from around Grade 3 onward with both being helpful for struggling readers. Crucially, Suggate estimated that only about 18% of such intervention studies provide long-term follow-up data, which did not provide for a sufficient sample size to systematically test the role of moderator variables, in particular intervention type. Accordingly, more research into the long-term effects of phonemic awareness, phonics, fluency, and comprehension reading interventions is needed—ideally via meta-analysis because this better accounts for Type II error and allows exploration of moderator variables (Hunter & Schmidt, 2004).

Intervention Features

In addition to it being important to examine the long-term effects of reading interventions, it may prove insightful to examine the influence of practical intervention features,

such as the necessary instructor–student ratio, duration of the intervention, and whether booster interventions help.

Instructor–student ratio. In the United States, the risk status of the readers often determines the intensity of reading intervention that they receive. For example, Tier III interventions typically targeting the lowest 10% of readers are often delivered with an instructor–student ratio of 1:1, whereas at-risk readers between the 10th and 25th percentiles typically receive small-group interventions (Scholin & Burns, 2012). Accordingly, in part to justify this practice, an important question to resolve is whether instructor–student ratio affects intervention outcomes, again with a long-term focus. To illustrate the point, if small group interventions were as effective as individual interventions, then resource allocation could be accordingly optimized by reaching a greater number of students with the same number of teachers. Generally, meta-analyses indicate no difference in effect size depending on whether instruction was delivered in small groups or individually (Ehri, Nunes, Willows, et al., 2001; Elbaum et al., 2000; Suggate, 2010; but cf. Ehri, Nunes, Stahl, et al., 2001); however, little is known about the effect of instructor–student ratio on long-term effect size.

Intervention administrator. Effective implementation of interventions depends not only on the content but also on the administrator (Marulis & Neuman, 2010). At a practical, financial level, the requisite qualifications for successful implementation of interventions is important. Internationally, a number of different intervention administrators have been employed, namely teaching assistants (Ryder, Tunmer, & Greaney, 2008) or paraeducators (Vadasy & Sanders, 2010), especially trained interveners (e.g., Center et al., 1995), regular classroom teachers (e.g., Elbaum et al., 2000), student peers (Fuchs & Fuchs, 2005), or computers (Cheung & Slavin, 2012). Finally, a well-documented phenomenon is that researcher-administered interventions tend to result in larger effect sizes (e.g., Dignath & Buttner, 2008; Ehri, Nunes, Willows, et al., 2001) and computer-led interventions smaller effects (Cheung & Slavin, 2012; Ehri, Nunes, Willows, et al., 2001). Conceptually, meta-analyses not taking sufficient account of intervention administrator run the risk of conflating administrator effects with intervention effects and accordingly run the risk of leading to false policy recommendations.

Intervention length and booster interventions. It is surprising that intervention length is seldom a significant predictor of intervention effect size in meta-analyses (e.g., Suggate, 2010). One possibility for this finding is that interventions have too narrow a focus, such that those focusing specifically on one domain, especially if the targeted skills are highly constrained (e.g., Paris, 2005), might be unable to drive further benefits beyond a certain point saturation

point. Conversely, many interventions now contain a mixed approach (see the appendix), and most include well-established outcome measures unlikely to be affected by ceiling effects. In addition, previous meta-analyses (e.g., Ehri, Nunes, Stahl, et al., 2001; Ehri, Nunes, Willows, et al., 2001; Suggate, 2010) have coded dosage of the intervention condition only, which ignores that in many studies, the control children receive an intervention, sometimes of the same duration (e.g., Antoniou & Souvignier, 2007; Fälth, Gustafson, Tjus, Heimann, & Svensson, 2013; Gunn, Smolkowski, & Vadasy, 2011). Thus, raw dosage of the intervention condition may be contaminated by not considering the dosage of the control group. Furthermore, some studies include booster interventions (Coyne, Kame'enui, Simmons, & Harn, 2004), which also need to be accounted for in estimating effect sizes.

Methodological Factors Affecting Effect Size

In addition to there being a lack of knowledge on the long-term effects of reading interventions with respect to intervention type, administrator, sample, and instructor–student ratio, we also lack understanding of possible methodological and conceptual moderator effects on intervention outcome at follow-up.

Methodological quality. In assessing the quality of (clinical) intervention studies, indices have been developed that focus on randomization, attrition, and blindness (Jadad et al., 1996). Although it is not possible to ensure administrator blindness while administering reading intervention, researchers' monitoring of treatment fidelity to some extent acts as a proxy because both ensure adherence to protocol. Furthermore, sample attrition and randomization are certainly key variables to consider in intervention research. Specifically, it is conceivable that attrition inflates effect size to follow-up as treatment nonresponders may opt out of the intervention. Equally plausible is that dissatisfied children in the control group seek out additional reading support, thus reducing effect size at follow-up; however, because assignment usually occurs at the class level in reading intervention studies, the latter case is less likely. With regard to experimental design, quasi-experimental designs have tended to produce greater effect sizes than studies employing random assignment (e.g., Cheung & Slavin, 2012; Suggate, 2010).

Moreover, it is crucial to consider sample size—not just for the calculation of weighted effect sizes—but because of the danger of publication bias (e.g., Hunter & Schmidt, 2004). Publication bias places a particular threat to follow-up investigations from two angles. First, it is highly unlikely that researchers of unsuccessful interventions at posttest would then invest the considerable time and effort required to conduct a follow-up assessment. Second, it seems

unlikely that researchers who do collect follow-up data but find that their intervention did not result in positive effects would be motivated to publish their work; and even if they were, given the difficulty in interpreting null findings in follow-up studies often suffering from high attrition rates, such work may not pass muster during the peer-review process.

Skill constraint. In addition, it has been suggested that some reading skills follow a more typical and short-lived learning curve trajectory, quickly reaching a ceiling in both their mastery and contribution to reading (Paris, 2005). Accordingly, it might be expected that improvements in more constrained skills—such as word decoding, alphabet knowledge, phonemic awareness—are easier to exert than on less-constrained measures of skills such as reading comprehension. Because more constrained skills can, by definition, exhibit lesser improvement, these might lead to smaller follow-up effect sizes, particularly as the control group subsequently make developmental gains postintervention. Therefore, one possibility that needs testing is whether the less-constrained skills of reading comprehension (Paris, 2005), reading of phonetically noncontrolled text, and spelling measures, thanks to English's irregular orthography (Seymour, Aro, Erskine, & COST Action Network, 2003), show greater follow-up effect sizes than more constrained alphabetic and decoding measures—particularly over the long term, as more time allows more children the opportunity to reach ceiling on constrained skills.

Transfer effects. It is crucial to understand the effects of reading interventions on long-term reading outcomes, not merely on constructs targeted by the intervention. It is expected that phonics interventions will exert large improvements on decoding and phonemic awareness measures, by virtue of these skills being finite and attainable with a low ceiling. It is not surprising that short-term effects of, for example, phonemic interventions indicate large effects on phonemic awareness outcomes (e.g., $d_w = 1.11$; Ehri, Nunes, Willows, et al., 2001). To understand transfer to broader reading skills, it is still unsatisfactory to look simply at short-term effects on nontargeted skills, because long-term, generalizable effects are sought. Thus, phonics interventions and comprehension interventions must show effects on reading and reading comprehension, not merely on targeted skills such as segmenting or comprehension strategy use. Again, definitive answering of this question requires examination of follow-up data.

Summary. Research is needed that examines the long-term effect sizes of reading interventions, particularly with respect to whether key variables such as intervention type, administration, and sample risk status play a role. Indeed, previous meta-analyses have reported long-term effects of phonemic awareness, phonics, fluency, and reading comprehension interventions, but these have tended to focus on

one intervention type (Bus & van Ijzendoorn, 1999; Ehri, Nunes, Stahl, et al., 2001; Ehri, Nunes, Willows, et al., 2001) or target sample (Suggate, 2010; Swanson et al., 1999), such that the number of studies reported has been too low to provide a reliable estimation. Moreover, such an analysis has the potential to test whether intervention and methodological characteristics play a role in determining successful intervention.

Current Study

This article reports the results of a meta-analytical test of experimental and quasi-experimental reading interventions focusing on phonemic awareness, phonics, fluency, and comprehension approaches that include a long-term follow-up postintervention. Needless to say, inclusion of only these interventions does not mean that other interventions are ineffective, but rather that the evidence base for short-term effects (Bus & van Ijzendoorn, 1999; Ehri, Nunes, Stahl, et al., 2001; Ehri, Nunes, Willows, et al., 2001; Suggate, 2010; Talbott et al., 1994) is sufficient to begin to investigate follow-up effects. In addition to coding intervention type, sample risk status and methodological and intervention features were coded to shed light on reasons for intervention effect size changes from posttest to follow-up.

Research Questions

Consistent with previous work, it was expected that the interventions would show positive short-term effect sizes that would decrease to follow-up. In addition to this hypothesis and given the availability of studies looking in a nuanced way at long-term effects, the following research questions were formulated.

1. What are the effect sizes for normal, at-risk, and low readers and reading disabled readers from posttest to follow-up?
2. To what degree do phonemic awareness, phonics, fluency, comprehension, and mixed interventions result in different effect sizes on different outcome measures (i.e., transfer effects)?
3. To what extent do sample characteristics, including grade, gender, and intervention language, relate to follow-up effect size?
4. How do the methodological quality indicators of sample attrition, experimental design, treatment fidelity, and sample size with respect to publication bias, influence effect size?
5. How do the intervention characteristics of intervention length and administrator (i.e., preschool teacher, trained intervener, computer, tutor, experimenter, class teacher), instructor–student ratio, months to follow-up, and the presence of a booster intervention relate to effect size?

Method

Procedure

Literature search and article screening. To reduce the likelihood of publication bias, both published peer-reviewed and nonpublished studies were considered. A three-tiered approach to searching for studies was taken. In the first, four sets of terms (listed below) were entered into both PsycINFO and ERIC. Within each set the OR command was used, and between sets the AND command was used to combine the data. The first search was conducted in 2010 and restricted to include articles published after 1980 and to samples up to Grade 7.

The search terms were, for Set 1, reading: reading, reading ability, reading fluency, reading strategies, reading achievement, oral reading, reading development, reading intervention, reading education, school-based intervention, phonics, phonemic awareness, reading comprehension, repeated reading, remedial reading, and reading recovery; for Set 2, design features: control group, reading-matched control, experimental, quasi-experimental, between-subjects, between-groups, randomized, randomized-control, design, random assignment, and treatment; for Set 3, metrics: reading measures, reading skills, reading speed, reading accuracy, effect size, academic achievement, vocabulary, grammar, syntax, and language; for Set 4, follow-up measurement: long-term, medium-term, follow-up, posttest, longitudinal, period, and maintained.

The PsycINFO and ERIC searches identified 557 and 508 articles, respectively, which had been written between 1980 and 2010. The abstracts of these articles were checked to see whether the article (a) was a reading intervention as here defined with a phonetic, decoding, comprehension, or fluency focus, (b) included a follow-up assessment, and (c) contained at least one control or comparison group. This narrowed the number of articles selected to 54 unique articles, once overlapping entries from the PsycINFO and ERIC searches were removed. Second, a search of the reference sections of four previous meta-analyses that utilized similar interventions as those investigated here was conducted (Bus & van Ijzendoorn, 1999; Ehri, Nunes, Stahl, et al., 2001; Ehri, Nunes, Willows, et al., 2001; Suggate, 2010; Wanzek et al., 2006). These, once the abstracts were checked, netted a further 21 articles selected for closer consideration.

To particularly thoroughly canvass recent studies conducted since the last published meta-analysis (i.e., Suggate, 2010) and also those published after the first search in 2008, a third search with expanded terms focused on the last 10 years (counting back from October 2013). An expanded set of terms were used to increase the probability of hits, namely, for Set 1: peer tutoring, peer assisted, phonics, strategy instruction, phonological awareness, early reading,

supplemental instruction, and fluency; for Set 2: computer assisted, computer, and instructional support; for Set 3, word reading, decoding, phonemic awareness, and spelling. The term *reading recovery* was removed from this search because, as alluded to by an anonymous reviewer, it was the only proper noun referring to a specific intervention. In ERIC, this search was limited to preschool, elementary school studies published in English, German, or French as journal articles, conferences papers, speeches/meetings, reports, dissertation theses, doctoral theses, master's theses, books, collected works, and proceedings, encompassing the time from 2010 to 2013.

This broader PsycINFO search identified 880 abstracts and the equivalent ERIC search, but restricted to the last 3 years resulted in 134 articles. Of these, the vast majority, 72.05% and 77.87%, respectively, were rejected out of hand because they did not contain a reading intervention design. Articles were examined more closely and coded by the author. During this process, of the remaining articles, most were dropped, the reasons for which were as follows: (a) studies did not have a posttest with follow-up design (61.28%), (b) did not contain a matched or randomly assigned control group (36.84%), (c) did not qualify as a reading intervention as here defined (7.14%), (d) were too old (5.64%), or (e) were published in a language other than English or German (1.25%). Further studies were excluded if there was insufficient statistical information and the authors could not be contacted, or the data analysis was conducted at a classroom level instead of at the individual student level or contained self-selected assignment to groups without reporting pretest scores. Only one study that was not published in a peer-reviewed journal potentially satisfied inclusion criteria but was excluded because it would not be possible with only one study to examine the independent variable of publication outlet. Therefore, this meta-analysis represents only peer-reviewed articles.

Article coding

Intervention outcome. Outcome variables were collected for prereading, reading, reading comprehension, and spelling measures, consistent with previous meta-analyses (Ehri, Nunes, Stahl, et al., 2001; Ehri, Nunes, Willows, et al., 2001; Suggate, 2010).

Prereading measures. Measures were classified as prereading if they targeted phonological or phoneme awareness (to the exception of word repetition measures), letter naming, sounding out letters or words, and pseudoword or phonetically controlled word reading. Such measures included the *Test of Phonological Processing* (Wagner, Torgesen, & Rashotte, 1999), *Phonemic Segmentation Fluency*, *Letter-Sound Identification*, *Letter Naming Fluency* (Good & Kaminski, 2003), and *Woodcock Word Attack* (Woodcock, 1998).

Reading measures. Reading measures were those that included word reading, when this was not confined to phonetically controlled words, passage reading, and reading accuracy scores. Examples include reading book level, the *Gray Oral Reading Test* accuracy scores (Wiederholt & Bryant, 1992), *Word Identification* (Woodcock, 1998), *Word Reading* (Clay, 1993), *Oral Reading Fluency*, and *Passage Reading Tests* (Good & Kaminski, 2003).

Reading comprehension. Tests targeting the comprehension of text, usually through questions, maze procedures, or retelling tasks, were included under reading comprehension. Tests included *Woodcock Passage Comprehension*, *Gray Oral Reading Test* comprehension (Wiederholt & Bryant, 1992), *Stanford Diagnostic Reading Test* comprehension (Stanford Achievement Test Series, 1990), *Gates-MacGinitie* comprehension (MacGinitie, 1978), and maze tests.

Spelling. Spelling tests were included if the words were not phonetically controlled, such that simple rules could not be applied, to ensure that this was not a constrained skill. Among spelling tests coded were the *Waddington Diagnostic Spelling Test* (Waddington, 1988), the *Wechsler Test of Reading Development* spelling, *Schonell* spelling, *Kaufman* spelling (Kaufman & Kaufman, 1985), and the *Wide Range Achievement Test* spelling (Jastak & Wilkinson, 1984).

Sample risk status and characteristics. Risk status was coded according to ecologically occurring categories in the intervention literature, which however also bore a close resemblance to the tier classification system adopted in many states (Scholin & Burns, 2012). A restrictive classification system was used, whereby sufficient evidence had to exist for samples to be classified in the next, more severe at-risk category (i.e., the starting point for classification was normality, not risk), thus providing a more conservative estimate of disability. Normal readers were those drawn directly from normal classrooms, whereas samples from a low socioeconomic status background (e.g., vast majority receiving free and reduced lunch), those reading below the 50th percentile, and those whose parents were diagnosed with dyslexia were classified as at risk. Children reading between the 11th and 25th percentiles were classified as low readers (approximately Tier II). Children below the 10th percentile or with a IQ-reading discrepancy of 1 standard deviation in the negative direction were classified as reading disabled (corresponding to Tier III). Finally, a category for learning disabled students was included, that is for students with a general learning disability other than dyslexia.

Sample grade, age, whether the sample was given the intervention in its native language, and whether the studies were carried out in an English-speaking country were also

coded. Finally, because it is commonly reported that boys are overrepresented in reading interventions, the percentage of study participants who were boys was calculated.

Intervention type. The presence of phonemic awareness, phonics, fluency, and comprehension components in the reading interventions was coded using the criteria published by Suggate (2010), whereby phonemic awareness (and phonological awareness) interventions focused on manipulation of sounds in the absence of text and phonics interventions focused on skill at reading connected text, to the exclusion of practice at reading sentences or single words (e.g., peer tutoring, repeated reading). Comprehension interventions were those that focused on strategies to decipher text and derive meaning without a phonics focus, such as summarizing, prior knowledge, and inferential thinking. Thus, reading recovery was here classified as a comprehension intervention.

In a second step, these components were recoded into interventions as follows: (a) phonemic awareness, if they only contained a PA component, (b) phonics, if they contained a phonics component with or without an additional PA component, (c) fluency, if they included a fluency component with or without phonics, (d) comprehension, if containing a comprehension component, with or without a fluency component, or (e) mixed, if containing comprehension and a phonics or phonemic awareness component. Therefore, these categories captured a pure language ability in phonemic awareness, a sound-symbol intervention in phonics and as predicted by the phonological linkage hypothesis (Hatcher et al., 1994), the role of practice and fluency building (Therrien, 2004), teaching reading comprehension skills (Wanzek et al., 2006), being the aim of reading (Gough & Tunmer, 1986), and mixed approaches.

Instructor variables. Intervention administrator was coded using dummy variables to allow for the possibility that interventions has more than one type of administrator. Similar but expanded criteria to Marulis and Neuman (2010) were used, resulting in (a) preschool teacher, (b) classroom teacher, (c) trained tutor including qualified teachers trained for the study, (d) researcher administered including the researcher's graduate students, (e) parent or home administered, (f) computer administered, or (g) peer or community reading partner. To accommodate international differences, preschool teachers were often kindergarten teachers in European countries, where, unlike in much of the United States, kindergarten is not part of regular school.

Length of instruction was estimated using two variables. The first was a calculation of the total number of intervention hours. Where precise information was not provided, the best estimation possible was calculated. For example, if an intervention length was given as 10 to 15 minutes a day, 5

days per week for 3 months, the length would be 12.5 minutes multiplied by 5 days multiplied by 12 weeks (which is slightly less than 3 months to allow for absences). The second dosage variable was a dummy variable for whether the intervention was replaced at an exact one-to-one ratio by a school or in-house intervention of similar quality. Finally, instructor–student ratio was also coded as was the number of months from posttest to follow-up.

Study methodology. Experimental design was coded as a dummy variable (1 = *random assignment*, 0 = *quasi-experimental*). Random assignment referred only to the random assignment of individual students to the intervention and control groups. For inclusion in the meta-analysis, quasi-experimental assignment of groups of students drawn from the same populations was necessary. This excluded designs comparing, for example, at-risk to normal readers. To enforce this criterion, pretest scores on quasi-experimental designs had to show equivalence ($d_w = \pm 0.50$ at pretest). Attrition was calculated by taking the number of students at follow-up divided by the number of students receiving the intervention around pretest. Consistent with recommendations (Jadad et al., 1996), fidelity was coded as a dummy variable, where fidelity of 1 indicated that the authors had made mention of attempts to monitor treatment adherence.

Interrater reliability. A graduate student in educational psychology independently coded the study characteristics for 16.33% of the studies. Cohen's kappa coefficients were calculated for the dichotomous variables of sample language, study language, random assignment, intervention administrator, instructor–student ratio, and intervention type, yielding a mean reliability estimate of .93. For the continuous variables of grade, attrition, percentage boys, and intervention length, reliability was estimated by dividing the number of increments (e.g., percentage, hours) agreed on, by those disagreed on summed with those agreed on. Reliability by this method was estimated as 90.83%. Initial disagreements were then discussed until a consensus was reached. On the important variable of intervention type, a second psychology graduate (master's) coded 18.31% of studies, obtaining 94.23% agreement. Following this initial calibration procedure, all studies were jointly coded a second time by the author and the psychology graduate with initial disagreements being resolved by mutual agreement.

Data Analysis

Effect size calculation. Effect sizes (d) were calculated by dividing the difference between the means by the pooled standard deviations (Hunter & Schmidt, 2004). Individual effect sizes were calculated for each of the measures reported by the authors, at pretest, posttest, and follow-up. Then, individual effect sizes were averaged into the categories of

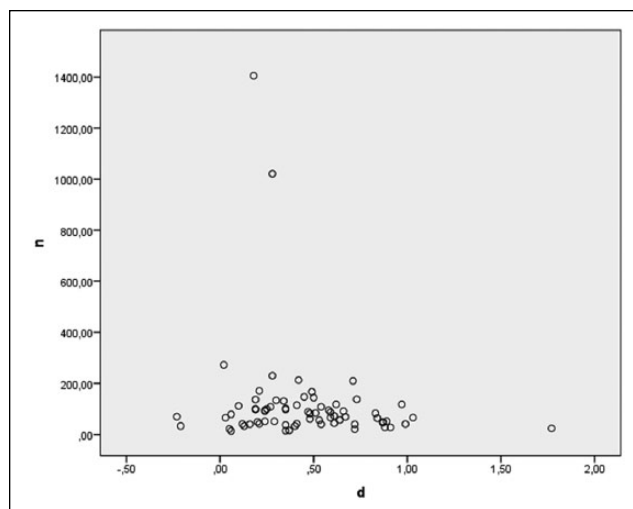


Figure 1. Funnel plot for effect size as a function of sample size at posttest.

prereading, reading, reading comprehension, and spelling. To retain statistical independence, effect sizes were never “counted twice,” in that they could feed into only one of the four categories. Once the four categories had been formed, an overall aggregate effect size was estimated by taking the mean of these four categories, consistent with previous meta-analyses (e.g., Ehri, Nunes, Stahl, et al., 2001; Ehri, Nunes, Willows, et al., 2001; Suggate, 2010). In a final step, the mean weighted effect sizes (d_w), also as a function of moderator variables, were calculated, with effect sizes being weighted by sample size, as recommended (Hunter & Schmidt, 2004). The Q -statistic was calculated ($p < .01$, given large number of comparisons) to estimate effect size heterogeneity and hence whether moderator variables likely operate. Fixed-effect effect size estimates were calculated.

Exploring bias. To first check for outliers, box plots were constructed, from which no data points at posttest or follow-up were identified as outliers ($> 2 SD$ above mean). To determine the presence of publication bias, funnel plots were constructed and these appear in Figure 1. There was a tendency for studies with smaller sample sizes and lower effect sizes to be absent at both time points, representing publication bias (Hunter & Schmidt, 2004). This pattern was most marked at follow-up with there appearing to be a corresponding mismatch around the median (i.e., which should be the midpoint of the funnel around which scores are mirrored), such that there were a lack of expected smaller studies with negative effect sizes.

Results

In all, 16 studies compared two or more different interventions with one control group. If the interventions were

coded as being of the same type, these groups were combined into a single intervention group; however, if the interventions were appreciably different, the sample size of the control group was divided by the number of intervention conditions to weight according to sample size (see Suggate, 2010). Where it was clear that there were large floor or ceiling effects, such that participants' scores were zero, or close to zero with a standard deviation larger than the mean, data for those particular measures were not coded (e.g., Brady, Fowler, Stone, & Winbury, 1994). In five instances the authors of original studies were contacted to clarify or obtain missing information (see Note 1). A further study (i.e., Blachman et al., 2014) reported, in addition to a 12-month follow-up, a 10-year follow-up; however, this was judged to be too great of an outlier with regard to follow-up to be included here, and instead the 12-month follow-up data are included. Only two studies had samples that had learning disabilities, so these were collapsed into the category of reading disabled, based on the reasoning that both represent severe learning impairment. A summary of the studies selected appears in the appendix. Of importance, the majority of the effect sizes at pretest (not reported) were at or close to zero, suggesting that the intervention and control groups were similar at the outset.

Study Descriptives

The mean time from posttest to follow-up was 11.17 months ($SD = 7.18$). In terms of risk status, 23.90% were classified as normal, 28.20% as at risk, 26.80% as low readers, and 21.10% as reading disabled. The interventions were predominantly administered by a mixture of teachers (21.10%), preschool teachers (16.90%), computers (21.10%), and trained educators (33.80%), with only 2.80% of interventions administered by parents and 7.00% by peer and community tutors. The majority of the interventions included either a phonemic awareness component (64.80%) or a phonics component (53.50%), whereas only 26.80% and 29.60% included components targeting fluency and comprehension, respectively. The average sample size was 125.25 ($SD = 211.23$) at pretest, 114.94 ($SD = 198.25$) at posttest, and 109.94 ($SD = 198.75$) at follow-up. Interventions lasted on average 38.70 hours ($SD = 37.13$), and these were usually conducted in English-speaking countries (60.60% of the time) and on participants in their mother tongue (87.30%) and employed random assignment and fidelity checks 52.10% and 54.90% of the time, respectively. The mean grade of the samples was Grade 1.18 ($SD = 1.51$), and on average 55.45% ($SD = 7.49$) of the intervention participants across studies were boys. The mean number of students per instructor was 4.89 ($SD = 6.81$), and the mean percentage of the pretest samples remaining at follow-up across studies (unweighted) was 83.91% ($SD = 18.62$). Authors reported that children received some form

of systematic intervention after posttesting in 12.70% of the studies.

Moderator Variable Analysis

Table 1 reports the mean weighted effect sizes, unweighted mean, median, sample size, number of treatment-control groups, estimated population standard deviation, and the Q test of effect size heterogeneity as a function of the categorical intervention moderators. Grade was grouped based on theoretical and power maximization criteria, resulting in the categories of preschool and kindergarten, then Grades 1 and 2, where children acquire decoding skills in English (Seymour et al., 2003), and then Grades 3 to 6, where children move to reading to learn (Chall, 1976) and which had too few studies to break students down further into individual grades.

Effect sizes in Table 1 tended to be of similar magnitude across outcome variables, with the exception of greater maintenance on spelling at follow-up. Normal readers appeared to lose their advantage over control groups to follow-up and experimenter administered interventions resulted in large effect sizes at posttest. Of interest, the long-term effect sizes were greater as a function of grade, such that kindergarten and preschool follow-up effect sizes were small, those in Grades 1 to 2 were small to moderate, and those in Grades 3 to 6 were large to moderate. Also, interventions that were conducted in addition to the control dosage showed a greater effect.

Next partial correlation analyses were conducted, controlling for sample size (e.g., Brannick, Yang, & Cafri, 2011; Hunter & Schmidt, 2004), to investigate the role of intervention length, sample attrition, time from posttest to follow-up, and grade. The resulting analysis appears in Table 2. Samples with a greater number of boys were associated with lower effect sizes at follow-up.

Intervention Type and Transfer Across Outcome Variables

Finally, to examine the role of intervention type in effect size and also the transfer of intervention effects to nontargeted outcomes, mean weighted effect sizes, unweighted mean, median, sample size, number of treatment control groups, estimated population standard deviation, and the Q test of effect size heterogeneity as a function intervention type were calculated. The resulting data, in Table 3, indicate that phonemic awareness and comprehension interventions resulted in the largest effect sizes at follow-up, whereas phonics and fluency interventions lost more effect to follow-up. Furthermore, fluency interventions did not result in good transfer to reading comprehension, with benefits being more confined to targeted decoding and reading skills at follow-up. Mixed interventions also showed generally sta-

Table 1. Effect Sizes at Posttest and Follow-Up as a Function of Outcome Measure, Risk Status, Administrator, Grade, Instructor Ratio, Experimental Design, Study and Participant Language, Booster Intervention, and Treatment Fidelity.

Comparison	Posttest							Follow-Up						
	d_w	d	Median	N	k	SD_{obs}	Q	d_w	d	Median	N	k	SD_{obs}	Q
Outcome measure														
Overall	.37	.45	.41	8,161	71	.19	139.70*	.22	.36	.33	7,806	71	.24	181.76*
Prereading	.39	.50	.40	5,693	57	.31	188.54*	.16	.33	.32	4,576	45	.29	138.97*
Reading skills	.32	.44	.46	6,005	48	.22	118.94*	.18	.37	.35	6,274	54	.30	190.90*
Comp.	.42	.41	.39	3,165	38	.11	46.89	.17	.27	.27	4,977	45	.25	120.15*
Spelling	.46	.42	.32	1,234	20	.08	21.65	.38	.43	.36	1,872	27	.16	38.41
Sample risk status														
Normal	.28	.40	.45	4,258	17	.19	53.42*	.13	.32	.36	4,174	17	.29	101.33*
At risk	.54	.64	.57	1,839	20	.23	42.88*	.35	.44	.41	1,650	20	.19	34.60
Low readers	.40	.40	.35	1,117	19	—	16.34	.32	.33	.27	1,073	19	—	12.49
Reading/learning disabled	.37	.32	.35	947	15	—	12.70	.30	.35	.33	909	15	.13	18.57
Intervention administrator														
Preschool teacher	.41	.49	.52	1,177	12	.20	23.70	.32	.48	.46	1,088	12	.20	22.14
Regular teacher	.28	.39	.28	3,801	15	.17	43.17*	.10	.28	.28	3,755	15	.28	86.68*
Trained intervener	.49	.51	.47	1,820	25	.13	32.64	.34	.36	.33	1,645	25	—	16.31
Computer	.31	.32	.35	709	14	—	11.45	.25	.31	.27	677	14	.15	17.72
Experimenter	.60	.66	.67	607	5	—	3.53	.36	.49	.43	604	5	.15	8.48
Peer/community tutor	.37	.42	.25	490	5	.34	18.87*	.30	.24	.28	475	5	.12	6.57
Grade														
Preschool and kindergarten	.34	.50	.51	3,817	29	.20	67.88*	.12	.36	.31	3,622	29	.33	125.25*
Grades 1 and 2	.40	.46	.41	3,115	30	.16	48.06	.26	.36	.36	3,000	30	.11	39.23
Grades 3 to 6	.35	.32	.27	1,229	12	.04	12.51	.43	.41	.36	1,184	12	—	11.76
Instructor ratio														
Individual	.38	.40	.36	1,735	30	.06	31.27	.28	.32	.27	1,659	30	.12	35.97
Group	.36	.49	.48	6,426	41	.22	118.61*	.20	.40	.39	6,147	41	.28	161.11*
Experimental design														
Quasi	.34	.46	.40	5,023	34	.19	77.97*	.18	.40	.34	4,816	34	.33	161.22*
Random	.41	.45	.41	3,138	37	.19	62.72*	.29	.33	.33	2,990	37	—	34.97
Study language														
English	.34	.44	.35	4,752	43	.19	84.62*	.17	.35	.33	4,535	43	.26	120.38*
Non-English	.40	.47	.49	3,409	28	.18	54.29*	.29	.38	.38	3,271	28	.20	58.52*
Booster														
Received booster	.50	.51	.49	800	9	.23	19.37	.20	.28	.15	738	9	—	8.00
No booster	.35	.44	.41	7,361	62	.18	117.12*	.22	.38	.38	7,068	62	.26	176.85*
Treatment fidelity														
Monitored	.32	.41	.35	5,378	39	.17	75.89*	.17	.35	.33	5,200	39	.27	131.81*
Not monitored	.45	.50	.47	2,783	32	.18	53.77*	.32	.38	.35	2,606	32	.14	44.05
Dosage														
Additional to control	.44	.47	.44	4,788	58	.17	89.71*	.31	.35	.33	4,446	58	.06	62.15
Same dosage as control	.26	.37	.28	3,373	13	.16	33.31*	.10	.42	.43	3,360	13	.40	145.09*

Note. d = unweighted average effect size; d_w = weighted estimated effect size; median = median effect size; SD_{obs} = observed SD; — = variance was (mathematically) negative due to second order sampling error.

* $p < .01$.

Table 2. Partial Correlation Coefficients for Continuous Hypothesized Moderator Variables Controlling for Sample Size.

	1	2	3	4	5	6	7
1. Age (k)	—	.33* (51)	.18 (39)	.29* (46)	-.28* (51)	-.14 (51)	.01 (51)
2. Attrition (k)		—	.33* (55)	-.04 (63)	-.11 (68)	-.15 (68)	-.22 (68)
3. % boys (k)			—	-.02 (53)	.12 (55)	-.23 (55)	-.31* (55)
4. Dosage (k)				—	.35* (63)	.01 (63)	-.09 (63)
5. Time to follow-up (k)					—	.11 (68)	-.15 (68)
6. Posttest d (k)						—	.40* (68)
7. Follow-up d							—

* $p < .05$.

Table 3. Effect Sizes at Posttest and Follow-Up as a Function of Intervention Type.

Intervention and Outcome	Posttest							Follow-Up						
	d_w	d	Median	N	k	SD_{obs}	Q	d_w	d	Median	N	k	SD_{obs}	Q
Phonemic awareness														
Overall	.43	.42	.42	1,156	17	—	11.19	.36	.46	.44	1,093	17	.07	18.31
Prereading	.40	.39	.40	760	13	—	7.15	.28	.43	.46	714	13	.19	19.15
Reading skills	.32	.31	.32	449	8	—	2.75	.33	.48	.53	634	10	.30	23.55*
Comp.	.32	.30	.20	346	6	—	4.83	.29	.28	.32	659	10	—	6.39
Spelling	.33	.20	.26	251	4	—	2.86	.38	.50	.36	716	9	.11	11.12
Phonics														
Overall	.29	.44	.43	4,162	22	.22	70.98*	.07	.25	.27	4,045	22	.25	82.04*
Prereading	.32	.48	.41	3,141	21	.27	77.90*	.08	.26	.25	3,024	21	.28	79.66*
Reading skills	.26	.45	.44	3,684	16	.26	76.32*	.07	.30	.27	3,895	20	.29	102.61*
Comp.	.47	.48	.53	351	4	—	1.22	-.10	-.03	-.04	1,952	8	.16	19.89*
Spelling	.38	.27	.31	223	5	—	2.99	.46	.48	.39	215	5	.14	5.60
Fluency														
Overall	.47	.59	.49	936	12	.22	23	.28	.33	.24	882	12	—	7.97
Prereading	.58	.80	.48	728	10	.55	61.26*	.42	.49	.46	300	4	—	3.62
Reading skills	.47	.56	.64	936	12	.22	22.37	.28	.31	.24	882	12	—	8.65
Comp.	.32	.45	.38	936	10	.18	17.42	.18	.21	.20	882	9	—	5.91
Spelling	.58	.65	.72	425	6	.16	8.47	.27	.30	.25	391	6	.13	7.64
Comp.														
Overall	.38	.33	.28	1,098	11	.14	16.43	.46	.42	.39	1,050	11	.10	13.27
Prereading	.41	.38	.18	261	4	.45	16.73*	.32	.22	.22	85	2	.38	4.92
Reading skills	.26	.18	.20	127	3	.34	6.37	.51	.46	.46	93	2	.38	5.08
Comp.	.38	.32	.26	964	10	.09	11.81	.46	.42	.39	1,050	11	.10	13.47
Spelling	.54	.50	.50	93	2	.28	3.71	.51	.46	.46	93	2	.38	5.08
Mixed														
Overall	.52	.51	.50	809	9	.17	14.65	.40	.44	.41	736	9	—	6.79
Prereading	.38	.35	.32	554	6	.19	10.81	.29	.29	.45	452	5	.27	12.70
Reading skills	.46	.46	.49	809	9	.21	17.15	.38	.45	.42	736	9	.08	10.01
Comp.	.52	.54	.50	722	8	.23	17.20	.39	.40	.33	599	7	—	1.69
Spelling	.44	.43	.31	241	3	—	2.36	.40	.39	.40	456	5	—	3.56

Note. d = unweighted average effect size; d_w = weighted estimated effect size; median = median effect size; SD_{obs} = observed SD; — = variance was (mathematically) negative due to second order sampling error.

* $p < .01$.

ble maintenance of effect size to follow-up across most outcome variables.

To facilitate interpretation of the results in Tables 1 and 3 in light of the advantage for older readers and comprehension interventions, descriptive statistics are reported for these variables. The mean grade of students receiving phonemic awareness ($M = 0.50$, $SD = 0.97$) and phonics interventions ($M = 0.45$, $SD = 0.81$) was in between kindergarten and Grade 1, whereas fluency ($M = 1.25$, $SD = 1.37$) and mixed interventions ($M = 1.61$, $SD = 0.70$) were given on average between Grades 1 and 2, and comprehension somewhat later, in Grade 3 ($M = 3.09$, $SD = 2.06$).

Discussion

A plethora of studies and even meta-analyses have documented the short-term effects of reading interventions for different learners (Bus & van Ijzendoorn, 1999; Ehri, Nunes, Stahl, et al., 2001; Ehri, Nunes, Willows, et al., 2001; Elbaum et al., 2000; Suggate, 2010; Swanson et al., 1999). Thanks to a large body of work encompassing single studies that report longer-term effects, this article could

present the first detailed investigation not only of the longer-term effects of reading interventions but also of these as a function of a host of key moderator variables.

Consistent with previous work (Bus & van Ijzendoorn, 1999; Ehri, Nunes, Stahl, et al., 2001; Ehri, Nunes, Willows, et al., 2001; Elbaum et al., 2000; Suggate, 2010; Swanson et al., 1999), effect sizes at posttest were moderate ($d_w = 0.37$), on average, suggesting that the children in the experimental groups did in fact experience a substantial boost to their reading skills, which reduced by follow-up to around $d_w = 0.22$ (see Table 1). Thus, on average, 11 months after participating in interventions with a phonemic awareness, phonics, fluency, or comprehension approach, a small effect of the intervention remained.

Moderator Effects

A key contribution of the current article is understanding the role of moderator variables. Beginning with methodological moderators, effect sizes at follow-up tended to be lower when the studies included treatment-fidelity monitoring or nonrandomized designs, had less attrition, had a more

even gender balance, and were not directly carried out by members of the research team. Of interest, normal readers appeared to profit least from reading intervention, especially at follow-up.

Grade. However, the key findings to arise from this meta-analysis did indeed emerge when the effect sizes at follow-up were examined with respect to two key moderator variables in particular, namely, intervention type and grade. Beginning with the latter, grade did not appear to moderate the short-term effects of reading interventions, unlike in previous work (e.g., Suggate, 2010); here it was particularly evident that the younger the intervention sample, the lower the effect size at follow-up, despite moderate posttest effect sizes. Thus, for preschoolers and kindergarteners (according to the U.S. usage of these terms), effect sizes reduced from $d_w = 0.34$ to $d_w = 0.12$ at follow-up; for children in Grades 1 and 2, the effect reduced from $d_w = 0.40$ at posttest to $d_w = 0.26$ at follow-up; for older children in Grades 3 to 4, this actually increased from $d_w = 0.35$ to $d_w = 0.43$ at follow-up. In other words, the reading interventions investigated here were more than 3.5 times more effective for older children at follow-up. On the surface, this would appear to run counter to the popular idea that if children are not caught early, they will learn a pattern of failure such that reading intervention will not be successful (Good, Simmons, & Smith, 1998) because reading intervention can be effective in the early grades, and can be particularly effective in the middle grades, showing stronger effects 1 year after cessation of the intervention.

Intervention type and the phonological linkage hypothesis. The second key finding to emerge from this meta-analysis concerns the performance of phonics interventions. According to the phonological linkage hypothesis (Hatcher et al., 1994), reading instruction that explicitly combines the links between sounds in letters or words (i.e., phonics) should be more effective than purely phonemic approaches. At immediate posttest, there was little evidence that it mattered whether or not phonics or purely phonemic awareness interventions were used ($d_w = 0.33$ vs. $d_w = 0.32$). However, when follow-up effect sizes were compared, there was a distinct advantage for phonemic awareness interventions ($d_w = 0.29$ vs. $d_w = 0.07$), precisely the opposite of what would be predicted by the phonological linkage hypothesis. Of interest, the greatest effect sizes at follow-up appeared to result from interventions with a comprehension component.

This fairly large effect for comprehension interventions must be tempered with the observation that older children tended to receive interventions with a comprehension component. Accordingly, due to the inability to statistically tease out the influence of grade from intervention type because of the cell sizes (there were only 12 studies after

Grade 3), it remains unclear what drives the larger effect sizes for older children. However, this question is somewhat irrelevant because comprehension interventions cannot be effectively given to children that cannot yet read (e.g., Suggate, 2010).

Indeed, the poor performance of phonics interventions in comparison to phonemic awareness training is a surprising finding of this analysis that is worthy of discussion, at the outset of which a number of potential explanations can be ruled out. First, this advantage for phonemic awareness does not run counter to previous research because this research has not tested, using meta-analyses with a sufficiently large sample size, the phonological linkage hypothesis by comparing phonics and phonemic awareness interventions at follow-up. Second, this finding is not due to some feature of the respective study participants because post hoc analyses revealed that both study sets had highly similar samples in terms of grade, risk status, attrition, gender, and time between posttest and follow-up. Third, findings cannot readily be explained in terms of the outcome measure selected because phonemic awareness showed a comparative and appreciable advantage over phonics on all outcome measures, except for spelling (i.e., transfer effects). Thus, given the large number of studies in each condition, it is unlikely that a methodological feature accounts for the advantage for phonemic awareness interventions over phonics.

Conversely, in meta-analysis the effect size is calculated as a function of a weight assigned to each study. Based on sampling error, studies with larger sample sizes are assigned greater weights. As recommended by Hunter and Schmidt (2004), this meta-analysis weighted each study according to the sample size of that study. This a priori decision was taken because of the intuitive appeal that such a parsimonious weighting system entails in the absence of the seemingly excessive data-transformation when weighting according to sampling error or inverse sampling variance. Of interest, Brannick et al. (2011) found in a simulation based on published meta-analysis data that weighting by sample size generally performed as well as or better than other weights (e.g., Brannick et al., 2011). However, a consequence of this weighting system is that larger studies are given greater weights than when inverse variance is used, for example. Moreover, the two largest sample sizes involved phonics interventions (Gunn et al., 2011, $N = 1,405$; Houtveen & van de Grift, 2012, $N = 1,021$), and they were considerably larger than the third largest ($N = 273$), but resulted in small effect sizes at follow-up ($d = -0.13$ and 0.12 , respectively), despite being moderate at posttest ($d = 0.18$ and 0.28 , respectively). Of these two studies, Gunn et al. (2011) clearly implemented a high-quality phonics program, as defined here. The Houtveen and van de Grift (2012) intervention components appeared to be phonics because the intervention feature that distinguished the

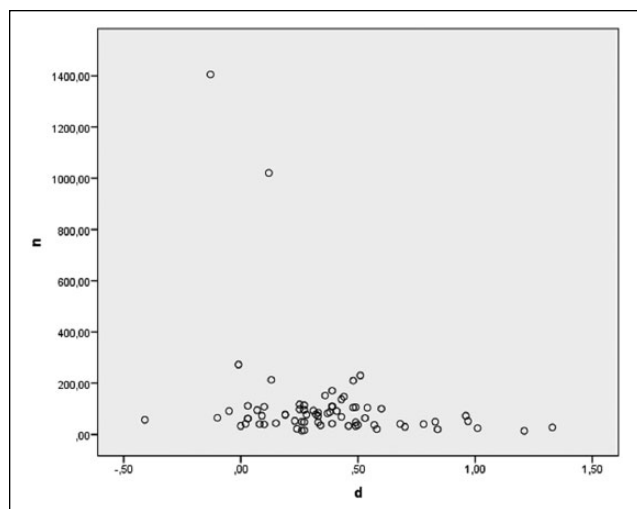


Figure 2. Funnel plot for effect size as a function of sample size at follow-up.

experimental from the control classes was the explicit instruction of letter–sound relations, although the methodology was less exactly described than in the Gunn et al. study. However, removing the Houtveen and van de Grift study minimally altered the follow-up effect sizes.

However, it would have been difficult to justify excluding either of these studies in the analyses or changing the meta-analytical methods simply because the findings may not fit with a theory such as the phonological linkage hypothesis (Hatcher et al., 1994). Both articles involved large-scale, real-world, quasi-experimental studies with some degree of randomization at the level of participating schools and treatment fidelity monitoring, and both resulted in positive short-term effects. Furthermore, given the evidence for publication bias among smaller studies found here, it would seem to be a strength of the current article that two large-scale studies were included to provide a pegging to the upper end of the funnel against which to compare smaller studies for publication bias (see Figures 1 and 2). Expressed more strongly, removing these studies would restrict our knowledge of reading interventions to smaller scale trials afflicted with publication bias.

Finally, phonics interventions showed heterogeneity in effect size, as indicated by the significant Q statistic at follow-up, whereas these were not generally significant for the other intervention types. Indeed, removing the Gunn et al. (2011) article reduced the heterogeneity, $Q(20) = 33.89, p = .03$, indicating that a moderator is operating in the estimation of phonics effect sizes. However, this does not answer the question of what this moderator is; it seems difficult to exclude the Gunn et al. article given that this article was a well-conducted phonics intervention. Perhaps instead this

points to the possibility that the moderator operating is some form of publication bias.

Conversely, an interesting question arises as to whether the inclusion of similar large-scale trials for comprehension and phonemic awareness interventions—had these existed—would also have accordingly reduced the obtained follow-up effect sizes for these as well. The answer to this question is purely hypothetical until such studies are conducted; however, insight might be gained from examining other large studies in the current analysis. After the two phonics articles in question, the next four largest studies (n in excess of 200 at follow-up in each case) contained one further phonics, two comprehension, and one phonemic awareness intervention. The effect sizes at follow-up for the phonemic awareness intervention were $d_w = 0.13$, for the third phonics intervention $d_w = 0.51$, and for the comprehension interventions $d_w = 0.51$ and 0.48 . Thus, the large comprehension interventions showed large effects, consistent with the weighted effect sizes reported in Table 2. Based on the small effect size for the phonemic awareness intervention, it might very tentatively be concluded that the provision of large-scale phonemic-awareness intervention studies would result in similar findings to those found for the phonics interventions.

Theoretically, it would appear difficult to explain why phonics interventions lost their effectiveness to follow-up so dramatically in comparison to other intervention types, especially in comparison with phonemic awareness. Perhaps the most simple explanation would be that all children—also including control group children—either receive instruction in phonics skills or develop these skills implicitly (Thompson et al., 2008) during regular schooling between posttest and follow-up, such that any advantage for the constrained phonics skills (Paris, 2005) washes out (Suggate, in press). However, this possibility would depend on children not receiving systematic instruction in phonemic awareness skills that also targets constrained skills, otherwise the advantage for phonemic awareness intervention at follow-up would not have been found here. Alternatively, the phonemic awareness interventions resulted in slightly greater short-term gains on prereading skills ($d_w = 0.40$ vs. $d_w = 0.32$); perhaps then the opposite to a washing out effect occurred, in that phonemic awareness resulted in a short-term advantage that escalated over time (Blachman et al., 2014), as predicted by Matthew effects (Pfof, Hattie, Dorfler, & Artelt, 2013; Stanovich, 1986).

Comprehension interventions. Although more, ideally large-scale, research is needed looking at phonemic and phonics interventions long term, the findings robustly indicate that comprehension interventions have good effects at follow-up on a host of skills not specifically targeted in the

interventions. There was some support for Paris's (2005) idea that skill constraint influences reading development. Thus, interventions targeting the least constrained construct, namely reading comprehension skills, exerted the greatest improvement to follow-up. Moreover, this improvement was not due to comprehension interventions showing improvement only on skills specifically targeted in the intervention, because skills typically targeted in reading comprehension training are generally higher-order meta-strategies, such as reflecting, summarizing, and predicting. However, reading comprehension tests measure understanding of text, which is in essence the goal of reading (Gough & Tunmer, 1986). In contrast, phonics ($d_w = -0.10$), but less so phonemic awareness ($d_w = 0.29$), measures showed lesser transfer to reading comprehension than comprehension or mixed interventions did ($d_w = 0.39$).

Intervention dosage. Of interest, as hypothesized, taking account of whether reading interventions were administered in addition to typical instruction instead of in place of it was related to effect size ($d_w = 0.44$ vs. $d_w = 0.26$ at posttest and $d_w = 0.31$ vs. $d_w = 0.19$ at follow-up). Specifically, the hours spent in reading intervention was not a significant variable here and in previous analyses, presumably because any amount of intervention has to occur outside of what children would otherwise receive to have an (appreciable) effect. Finally, offering a booster intervention was associated with lower effect sizes. One possibility for this counterintuitive finding is that more of the same may not work—perhaps booster interventions should contain a different approach to the mother intervention. However, the most likely explanation is perhaps that booster interventions were offered to students who either regressed or were not showing the desired progress (i.e., they were treatment resisters; Coyne et al., 2004), hence making the disadvantage for booster interventions a product of particular samples, not the general practice of giving booster interventions.

Limitations

Overall, the funnel plots suggested a publication bias (Hox, 2010), as evidenced by a lack of small to medium-sized studies. Accordingly, many of the effect sizes should be interpreted with caution, as they are likely slightly upwardly biased. Although, as discussed, the inclusion of the two larger studies that showed modest effects would probably have compensated for publication bias, this is clearly speculative until other very large-scale trials are conducted. Moreover, studies with long-term follow-ups that were not published in peer-reviewed journals were not sufficiently methodologically rigorous to include here. Even if methodologically rigorous unpublished studies

were to be found, this would not solve the bias problem—as it is highly likely that researchers of interventions that do not result in positive and appreciable short-term effects would not collect follow-up data. Perhaps the only way to circumvent this problem would be to establish a database of educational intervention studies, whereby researchers register their studies before commencing them, so that the proportion of studies making it to a follow-up data collection can be estimated against the number of “failed” studies.

Ceiling effects in the data did not appear to have constrained the reading development of children at follow-up, thus suppressing effect size. All studies were of high quality, and the means and standard deviations were inspected for possible ceiling effects during the coding. Even the pre-reading measures considered to be constrained, according to the rationale provided by Paris (2005), did not seem to show evidence of ceiling effects based on means and standard deviations reported in studies. An informal observation was that authors tended to developmentally shift the tasks that children received, such that a measure of letter-sound correspondences in kindergarten pretest was not readministered at follow-up, but was instead replaced with a more difficult and less constrained (for the given age group) measure of nonword reading, for example.

Many studies did not provide information on the kinds of reading experiences children received after the intervention, such that it was presumably not possible to entirely reliably code whether intervention or control groups received a booster intervention, which tempers the findings concerning booster effects. In some instances it is likely that the children originally assigned to the control group underwent subsequent reading intervention given by the school, possibly in greater numbers than in the treatment groups, but which was not reported by the authors. These posttest interventions may have resulted in control group gains, reducing the effect size advantage for the treatment group at follow-up; to solve this problem, study authors are encouraged to more systematically collect data on posttest experiences of both samples. Finally, given the small number of studies, it was not possible to conduct regression analyses to try to tease apart the influence of age and comprehension interventions, in particular.

Implications

A key argument of the current article has been that long-term effects are key. This assumption is entirely justifiable in that neither the resources nor the desire exists to have children repeatedly and continuously in reading intervention. On the other hand, long-term effects are not the only criterion in evaluating effectiveness because in the absence of short-term effects, the effort of participating and

conducting reading interventions would unlikely be sufficiently rewarding to ensure necessary engagement.

Overall, the effect sizes obtained for reading interventions after, on average, 11 months are appreciable, but disappointing, particularly in light of a likely necessary downward adjustment due to publication bias. Indeed, the current effect size at follow-up of $d_w = 0.22$, pending a possible downward adjustment, is somewhat low, and for kindergarten and preschool children, the obtained estimate of $d_w = 0.12$ can be considered marginal. Of interest, such a small effect size among children younger than Grade 1 is consistent with quasi-experimental work suggesting that the effects of such early interventions tend to wash out (Dollase, 2007; Durkin, 1974–1975; Elley, 1992; Schmerkotte, 1978; Suggate, 2009, in press; Suggate, Schaughency, & Reese, 2013). On the more positive side, there was evidence to suggest that reading interventions can be particularly effective for older readers ($d_w = 0.43$) and for reading comprehension interventions ($d_w = 0.47$). Of interest, both of these findings are compatible with what has been named the Luke effect, whereby reading instruction is predicted to have a greater effect when able to target less-constrained skills and when children have a greater skill base to draw on Suggate (in press).

In terms of the long-term effects of intervention type and its likely transfer to the broader construct of reading comprehension, the current analysis would suggest that preschool and kindergarten interventions should target phonemic awareness alone, leaving decoding skills to Grades 1 and 2. In Grades 1 and 2, fluency and mixed interventions appear to be optimal, although in the case of the former, effects may not transfer so well to reading comprehension. From Grade 3 onward, reading comprehension interventions would appear, on average, optimal. This is not to discount the importance of tailoring interventions to individual child needs (Connor, Morrison, & Katch, 2004; Connor et al., 2009); however, some clear indications were provided from this meta-analysis that may not have been uncovered by examining only the short-term effects.

However, before drawing conclusions about the effectiveness of reading intervention, the current analysis found indications of ways in which reading intervention studies can be tightened to better estimate effects. To recapitulate, effect sizes depended on important features of designs, namely whether studies employed random assignment and treatment fidelity monitoring, included samples better matched on gender, and contained larger sample sizes. Indeed, it was the author's subjective impression that more

recently published studies tended to be more comprehensively described and conducted. This, in turn, may explain why the effect sizes obtained here were a little lower than those of some previous analyses (Bus & van Ijzendoorn, 1999; Ehri, Nunes, Stahl, et al., 2001; Ehri, Nunes, Willows, et al., 2001; Suggate, 2010).

A key and unique finding from this meta-analysis is the greater retention of intervention effect to follow-up for at-risk, low, and disabled readers in comparison to normal readers. This findings is certainly encouraging for interventionists targeting struggling readers, suggesting that promising long-term effects are attainable. There was no reliable indication that one-to-one interventions were associated with greater effect sizes, perhaps calling into question whether Tier III students (similar to the reading disabled category used here) are always best treated by one-to-one interventions (see also Scholin & Burns, 2012). Based on the current study, it would appear more important that students in need receive the appropriate services, with it being less important if these are offered in individual or small-group settings. It is encouraging that the findings suggest that intervention administrator is, by and large, not the key determinant of effect size, with the exception that experimenters tended to exact larger short-term effect sizes out of their interventions and classroom teachers smaller effect sizes. Perhaps a more important factor for future research would be to examine the interaction between teacher qualification and treatment fidelity, to test whether less qualified teachers need to adhere more closely to treatment protocol to exact the same effects.

Conclusions

In conclusion, this meta-analysis extends our understanding of the effectiveness of reading interventions by providing a detailed analysis of the long-term effects. Indeed, in doing so, some surprising findings emerged, namely that phonemic awareness interventions appeared better than phonics, which is inconsistent with the phonological linkage hypothesis (Hatcher et al., 2004). Comprehension interventions, on the other hand, appeared particularly effective, as did those given to older pupils. Perhaps the greater contribution of this meta-analysis might be the challenges it lays down for researchers and journal editors to continue to improve on the quality of published studies and to conduct more large-scale follow-up investigations. It is hoped that this article will help stimulate research in this direction.

Appendix

Study Characteristics and Effect Sizes

Authors	Year	Type	Intervention Administrator	Grade	Risk Status	Posttest		Follow-Up		Months
						<i>d</i>	<i>n</i>	<i>d</i>	<i>n</i>	
Antoniou and Souvignier	2007	Comp.	Class teacher	6.5	RD	0.61	73	0.96	73	3
Blachman et al.	2004	Fluency	Researcher + trained admin.	2.5	Low readers	0.67	69	0.43	69	12
Blachman et al.	1999	Phonics	Preschool teachers	K	At risk	0.73	138	0.49	106	12
Brady et al.	1994	PA	Class teacher	K	At risk	0.21	42	0.68	41	6
Burns et al.	2008	Fluency	Peer/community tutor	5.0	Low readers	0.25	100	0.60	100	12
Büyüktaşkapu	2012	Mixed	Parent administered	1.0	At risk	0.87	50	0.83	50	4
Center et al.	1995	Comp.	Trained intervener	K	Low readers	0.89	52	0.97	52	12
Cirino et al. (English)	2009	Mixed	Trained intervener	1.0	At risk	0.34	131	0.39	111	4
Cirino et al. (Spanish)	2009	Mixed	Trained intervener	1.0	At risk	0.50	144	0.54	104	4
Clarke et al.	2010	Comp.	Trained intervener	3.0	Low readers	0.29	52	0.27	49	4
Clarke et al.	2010	Comp.	Trained intervener	3.0	Low readers	0.20	50	0.49	48	4
Clarke et al.	2010	Comp.	Trained intervener	3.0	Low readers	0.24	52	0.26	50	4
Dion et al.	2010	Phonics	Class teacher + peer/community tutor	K	At risk	0.83	83	0.28	77	12
Dion et al.	2010	Phonics	Class teacher + peer/community tutor	K	At risk	1.03	66	0.03	62	12
Dion et al.	2010	Phonics	Class teacher + peer/community tutor	K	Normal	-0.23	70	-0.10	65	12
Ecalte et al.	2009	Phonics	Computer	1.0	At risk	0.88	28	1.33	28	9
Elbro & Peterson	2004	Phonics	Preschool teachers	K	At risk	0.48	82	0.37	82	4
Fälth et al. comp	2013	Comp.	Computer	2.0	RD	-0.21	33	0.00	33	9
Fälth et al. mixed	2013	Mixed	Computer	2.0	RD	0.13	33	0.49	33	9
Fälth et al. phonics	2013	PA	Computer	2.0	RD	0.40	33	0.46	33	9
Fawcett et al.	2001	Mixed	Researcher	2.0	RD	0.59	87	0.38	87	6
Gittelman & Feingold	1983	Phonics	Trained intervener		RD	0.53	56	0.33	48	3
Gunn et al. (Hispanic)	2005	Fluency	Trained intervener	1.0	At risk	0.41	115	0.25	117	24
Gunn et al. (non-Hispanic)	2005	Fluency	Trained intervener	1.0	At risk	0.24	95	0.19	77	24
Gunn et al.	2011	Phonics	Class teacher	K	Normal	0.18	1405	-0.13	1405	12
Hatcher et al.	2004	Phonics	Preschool teachers	K	At risk	0.19	137	0.43	137	8
Hatcher et al.	2004	Phonics	Preschool teachers	K	Normal	0.02	273	-0.01	273	8
Hatcher et al.	1994	PA	Trained intervener	2.0	RD	0.16	40	0.08	40	9
Hatcher et al.	1994	Mixed	Trained intervener	2.0	RD	0.41	42	0.39	42	9
Hatcher et al.	1994	Comp.	Trained intervener	2.0	RD	0.12	41	0.02	41	9
Hook et al.	2001	PA	Computer		Low readers	0.05	22	0.24	22	20
Houtveen & van de Grift	2012	Phonics	Class teacher	1.0	Normal	0.28	1021	0.12	1021	12
Kjeldsen et al. high dose	2003	PA	Preschool teachers	K	Normal	0.49	167	0.36	152	25
Kjeldsen et al. low dose	2003	PA	Preschool teachers	K	Normal	0.99	41	0.10	39	25
Kozminsky & Kozminsky	1995	PA	Preschool teachers + trained admin.	K	Normal	0.48	61	0.70	30	9
Kyle et al.	2013	Phonics	Computer	2.0	Low readers	0.35	15	0.26	15	4
Kyle et al.	2013	Phonics	Computer	2.0	Low readers	0.37	16	0.27	16	4
Lie	1991	PA	Class teacher	1.0	Normal	0.45	147	0.44	147	12
Loeb et al.	2009	PA	Computer	2.0	Low readers	0.03	66	0.03	62	6
Loeb et al.	2009	PA	Computer	2.0	Low readers	0.35	38	0.34	36	6
Lyster	2002	Phonics	Preschool teachers	Pre	Normal	0.62	118	0.27	114	14
Mantzicopoulos et al.	1992	Fluency	Trained intervener	1.0	At risk	0.27	108	0.10	108	12
Morris et al.	2012	Phonics	Class teacher	1.5	RD	0.24	92	-0.05	92	12
Morris et al.	2012	Mixed	Class teacher	1.5	RD	0.66	92	0.41	92	12
Morris et al.	2012	Phonics	Class teacher	1.5	RD	0.58	96	0.27	96	12
Nancollis et al.	2005	PA	Researcher	K	At risk	0.42	213	0.13	213	24
O'Connor et al.	1998	PA	Preschool teachers	K	LD	0.06	14	1.21	14	12
O'Connor et al.	1998	PA	Preschool teachers	K	Normal	0.59	66	0.53	64	12
Phillips et al.	1996	Comp.	Parent administered	K	At risk	0.30	134	0.31	93	48
Regtvoort & van der Leij	2007	Phonics	Computer	K	At risk	0.64	57	-0.41	57	5
Reitsma & Wesseling	1998	PA	Computer	K	Normal	0.35	98	0.25	98	4
Rothe et al. K	2004	PA	Preschool teachers	K	Normal	0.54	40	0.78	40	6
Rothe et al. preschool	2004	PA	Preschool teachers	Pre	Normal	0.72	40	0.50	37	6
Ryder et al.	2008	Fluency	Trained intervener	1.5	At risk	1.77	24	0.84	20	18
Schachter & Jo	2005	Mixed	Trained intervener	1.0	At risk	0.97	118	0.48	105	10
Segers & Verhoeven	2005	Phonics	Computer	K	Normal	0.19	100	0.32	78	4
Spörer et al.	2009	Comp.	Researcher	4.5	Normal	0.71	210	0.48	210	3
Torgesen et al.	2010	Phonics	Computer + trained admin.	1.0	Low readers	0.54	108	0.39	108	12
Vadasy & Sanders LM	2013	Fluency	Trained intervener	1.0	Low readers	0.19	98	0.07	95	24
Vadasy & Sanders LM	2012	Fluency	Trained intervener	K	Low readers	0.51	84	0.19	77	24

(continued)

Appendix (continued)

Authors	Year	Type	Intervention Administrator	Grade	Risk Status	Posttest		Follow-Up		Months
						<i>d</i>	<i>n</i>	<i>d</i>	<i>n</i>	
Vadasy & Sanders non-LM	2013	Fluency	Trained intervener	1.0	Low readers	0.47	89	0.33	85	24
Vadasy & Sanders non-LM	2012	Fluency	Trained intervener	K	Low readers	0.84	64	0.23	53	24
Vadasy et al.	2006	Fluency	Trained intervener	K	Low readers	0.61	44	0.15	44	12
Vadasy et al.	2000	Fluency	Trained intervener	1.0	At risk	0.87	46	0.57	37	12
van der Kooy-Hofland et al.	2012	Phonics	Computer	K	Low readers	0.06	79	0.09	73	8
van der Kooy-Hofland et al.	2012	Phonics	Computer	K	Low readers	0.72	21	0.58	21	8
van Keer STRAT	2004	Comp.	Class teacher	5.0	Normal	0.28	231	0.51	231	4
van Keer STRAT + tutoring	2004	Comp.	Class teacher + peer/community tutor	5.0	Normal	0.21	171	0.39	171	4
Warrick et al.	1993	PA	Researcher	K	At risk	0.91	28	1.01	25	12
Williams	1980	Phonics	Class teacher		LD	0.35	102	0.33	72	7
Wolff	2011	Mixed	Class teacher	3.0	RD	0.10	112	0.03	112	12

Note. Comp. = reading comprehension; K = kindergarten; LD = learning disabled; LM = language minority; PA = phonemic awareness; pre = preschool; RD = reading disabled. In instances where samples were divided among groups, means were rounded.

Acknowledgments

Part of this work was completed while the author was receiving a fellowship from the Alexander-von-Humboldt foundation, hosted by Wolfgang Schneider, at the University of Würzburg. I also thank Elisabeth Neudecker and Tamara Suggate for their help with data coding and the latter also for conceptual assistance.

Declaration of Conflicting Interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

The author(s) received no financial support for the research, authorship, and/or publication of this article.

Note

1. One further study met inclusion criteria but had missing data that could not be retrieved due to damage to the building arising from the Christchurch earthquake in February 2011.

References

References marked with an asterisk indicate studies included in the meta-analysis.

- *Antoniou, F., & Souvignier, E. (2007). Strategy instruction in reading comprehension: An intervention study for students with learning disabilities. *Learning Disabilities: A Contemporary Journal*, 5, 41–57.
- Bell, S. M., McCallum, S., & Cox, E. A. (2003). Toward a research-based assessment of dyslexia: Using cognitive measures to identify reading disabilities. *Journal of Learning Disabilities*, 36, 505–516.

- *Blachman, B. A., Ball, E. W., Black, R. S., & Tangel, D. M. (1994). Kindergarten teachers develop phoneme awareness in low-income, inner-city classrooms. *Reading and Writing: An Interdisciplinary Journal*, 6, 1–18.
- *Blachman, B. A., Schatschneider, C., Fletcher, J. M., Francis, D. J., Clonan, S. M., Shaywitz, B. A., & Shaywitz, S. E. (2004). Effects of intensive reading remediation for second and third graders and a 1-year follow-up. *Journal of Educational Psychology*, 96, 444–461.
- Blachman, B. A., Schatschneider, C., Fletcher, J. M., Murray, M. S., Munger, K. A., & Vaughn, M. G. (2014). Intensive reading remediation in grade 2 or 3: Are there effects a decade later? *Journal of Educational Psychology*, 106, 46–57. doi:10.1037/a0033663
- *Blachman, B. A., Tangel, D. M., Ball, E. W., Black, R., & McGraw, C. K. (1999). Developing phonological awareness and word recognition skills: A two year intervention with low-income, inner-city children. *Reading and Writing: An Interdisciplinary Journal*, 11, 239–273.
- *Brady, S., Fowler, A., Stone, A., & Winbury, N. (1994). Training phonological awareness: A study with inner-city kindergarten children. *Annals of Dyslexia*, 44, 26–59.
- *Brannick, M. T., Yang, L., & Cafri, G. (2011). Comparison of weights for meta-analysis or *r* and *d* under realistic conditions. *Organizational Research Methods*, 14, 587–607. doi:10.1177/1094428110368725
- Buckingham, J., Wheldall, K., & Beaman, R. (2012). A randomised control trial of a Tier-2 small-group intervention (“MiniLit”) for young struggling readers. *Australian Journal of Learning Difficulties*, 17, 79–99. doi:10.1080/19404158.2012.717537
- *Burns, M. K., Senesac, B. J., & Silbergliitt, B. (2008). Longitudinal effect of a volunteer tutoring program on reading skills of students identified as at-risk for reading failure: A two-year follow-up study. *Literacy Research and Instruction*, 47, 27–37.
- Bus, A. G., & van Ijzendoorn, M. H. (1999). Phonological awareness and early reading: A meta-analysis of experimental training studies. *Journal of Educational Psychology*, 91, 403–414.

- *Büyüktaşkapu, S. (2012). Effect of family supported pre-reading training program given to children in preschool education period on reading success in primary school. *Educational Sciences: Theory & Practice*, *12*, 309–316.
- *Center, Y., Wheldall, K., Freeman, L., Outhred, L., & McNaught, M. (1995). An evaluation of reading recovery. *Reading Research Quarterly*, *30*, 240–263.
- Chall, J. (1976, April). *The great debate: Ten years later, with a modest proposal for reading stages*. Paper presented at the Conference on Theory and Practice of Beginning Reading Instruction, Pittsburgh, PA. Retrieved from <http://eric.ed.gov/PDFS/ED155617.pdf>
- Cheung, A. C., & Slavin, R. E. (2012). How features of educational technology applications affect student reading outcomes: A meta-analysis. *Educational Research Review*, *7*(3), 198–215. doi:10.1016/j.edurev.2012.05.002
- *Cirino, P. T., Vaughn, S., Linan-Thompson, S., Cardenas-Hagan, E., Fletcher, J. M., & Francis, D. J. (2009). One-year follow-up outcomes of Spanish and English interventions for English language learners at risk for reading problems. *American Educational Research Journal*, *46*, 744–781.
- *Clarke, P. J., Snowling, M. J., Truelove, E., & Hulme, C. (2010). Ameliorating children's reading comprehension difficulties: A randomized controlled trial. *Psychological Science*, *21*, 1106–1116. doi:10.1177/0956797610375449
- Clay, M. M. (1993). *An observation survey of early literacy achievement*. Auckland, New Zealand: Heinemann.
- Coltheart, M., Curtis, B., Atkins, P., & Haller, M. (1993). Models of reading aloud: Dual-route and parallel-distributed-processing approaches. *Psychological Review*, *100*, 589–608.
- Connor, C. M., Morrison, F. J., & Katch, L. E. (2004). Beyond the reading wars: Exploring the effect of child-instruction interactions on growth in early reading. *Scientific Studies of Reading*, *8*, 305–336. doi:10.1207/s1532799xssr0804_1
- Connor, C. M., Piasta, S. B., Fishman, B., Glasney, S., Schnatscheider, C., Crowe, E., . . . Morrison, F. J. (2009). Individualizing student instruction precisely: Child × instruction interactions on first graders' literacy development. *Child Development*, *80*, 77–100.
- Coyne, M. D., Kame'enui, E., Simmons, D. C., & Harn, B. A. (2004). Beginning reading intervention as inoculation or insulin: First-grade reading performance of strong responders to kindergarten intervention. *Journal of Learning Disabilities*, *37*, 90–104.
- Dignath, C., & Buttner, G. (2008). Components of fostering self-regulated learning among students. A meta-analysis on intervention studies at primary and secondary school level. *Metacognition and Learning*, *3*, 231–264. doi:10.1007/s11409-008-9029-x
- *Dion, E., Brodeur, M., Gosselin, C., Campeau, M., & Fuchs, D. (2010). Implementing research-based instruction to prevent reading problems among low-income students: Is earlier better? *Learning Disabilities Research & Practice*, *25*, 87–96.
- Dollase, R. (2007). Bildung im Kindergarten und Früheinschulung: Ein Fall von Ignoranz und Forschungsamnesie [Learning in kindergarten and preschool: A case of ignorance and forgotten results]. *Zeitschrift für Pädagogische Psychologie*, *21*, 5–10.
- Durkin, D. (1974–1975). A six year study of children who learned to read in school at the age of four. *Reading Research Quarterly*, *10*, 9–61.
- *Ecalte, J., Magnan, A., & Calmus, C. (2009). Lasting effects on literacy skills with a computer-assisted learning using syllabic units in low progress readers. *Computers and Education*, *52*, 554–561.
- Ehri, L. C., Nunes, S. R., Stahl, S. A., & Willows, D. M. (2001). Systematic phonics instruction helps students learn to read: Evidence from the National Reading Panel's meta-analysis. *Review of Educational Research*, *71*, 393–447.
- Ehri, L. C., Nunes, S. R., Willows, D. M., Schuster, B. V., Yaghouh-Zadeh, Z., & Shanahan, T. (2001). Phonemic awareness instruction helps children learn to read: Evidence from the National Reading Panel's meta-analysis. *Reading Research Quarterly*, *36*, 250–287.
- Elbaum, B., Vaughn, S., Hughes, M. T., & Moody, S. W. (2000). How effective are one-to-one tutoring programs in reading for elementary students at risk for reading failure? A meta-analysis of the intervention research. *Journal of Educational Research*, *92*, 605–619.
- *Elbro, C., & Peterson, D. K. (2004). Long-term effects of phoneme awareness and letter sound training: An intervention study with children at risk for dyslexia. *Journal of Educational Psychology*, *96*, 660–670.
- Elley, W. B. (1992). *How in the world do students read? IEA study of reading literacy*. The Hague, Netherlands: International Association for the Evaluation of Educational Achievement.
- *Fälth, L., Gustafson, S., Tjus, T., Heimann, M., & Svensson, I. (2013). Computer-assisted interventions targeting reading skills of children with reading disabilities—A longitudinal study. *Dyslexia*, *19*, 37–53.
- *Fawcett, A. J., Nicolson, R. I., Moss, H., & Nicolson, M. K. (2001). Effective reading intervention in junior school. *Educational Psychology*, *21*, 299–312.
- Ferrer, E., Shaywitz, B. A., Holahan, J. M., Marchione, K., & Shaywitz, S. E. (2010). Uncoupling of reading and IQ over time: Empirical evidence for a definition of dyslexia. *Psychological Science*, *21*, 93–101.
- Fuchs, D., & Fuchs, L. S. (2005). Peer-assisted learning strategies: Promoting word recognition, fluency, and reading comprehension in young children. *Journal of Special Education*, *39*, 34–44.
- *Gittelman, R., & Feingold, I. (1983). Children with reading disorders: Efficacy of reading remediation. *Journal of Child Psychology & Psychiatry*, *24*, 167–191.
- Good, R. H., & Kaminski, R. A. (2003). *Dynamic Indicators of Basic Early Literacy Skills: Administration and scoring guide* (6th ed.). Longmont, CA: Sopris West.
- Good, R. H., Simmons, D. C., & Smith, S. B. (1998). Effective academic interventions in the United States: Evaluating and enhancing the acquisition of early reading skills. *School Psychology Review*, *27*, 45–56.
- Gough, P. B., & Tunmer, W. E. (1986). Decoding, reading and reading disability. *Remedial & Special Education*, *7*, 6–10.
- *Gunn, B., Smolkowski, K., Biglan, A., Black, C., & Blair, J. (2005). Fostering the development of reading skill through

- supplemental instruction: Results for Hispanic and non-Hispanic students. *Journal of Special Education*, 39, 66–85.
- *Gunn, B., Smolkowski, K., & Vadasy, P. (2011). Evaluating the effectiveness of Read Well Kindergarten. *Journal of Research on Educational Effectiveness*, 4, 53–86.
- Harm, M. W., & Seidenberg, M. S. (1999). Phonology, reading acquisition, and dyslexia: Insights from connectionist models. *Psychological Review*, 106, 491–528.
- *Hatcher, P. J., Hulme, C., & Ellis, A. W. (1994). Ameliorating early reading failure by integrating the teaching of reading and phonological skills: The phonological linkage hypothesis. *Child Development*, 65, 41–57.
- *Hatcher, P. J., Hulme, C., & Snowling, M. J. (2004). Explicit phoneme training combined with phonic reading instruction helps young children at risk of reading failure. *Journal of Child Psychiatry*, 45, 338–358.
- *Hook, P. E., Macaruso, P., & Jones, S. (2001). Efficacy of fast forward training on facilitating acquisition of reading skills by children with reading difficulties—A longitudinal study. *Annals of Dyslexia*, 51, 75–96.
- *Houtveen, T., & van de Grift, W. (2012). Improving reading achievements of struggling learners. *School Effectiveness and School Improvement*, 23(1), 71–93. doi:10.1080/09243453.2011.600534
- Hox, J. J. (2010). *Multilevel analysis: Techniques and applications*. New York, NY: Routledge.
- Hunter, J. E., & Schmidt, F. L. (2004). *Methods of meta-analysis: Correcting error and bias in research findings* (2nd ed.). Thousand Oaks, CA: Sage.
- Jadad, A. R., Moore, A., Carroll, D., Jenkinson, C., Reynolds, D. J. M., Gavaghan, D. J., & McQuay, H. J. (1996). Assessing the quality of reports of randomized clinical trials: Is blinding necessary? *Controlled Clinical Trials*, 17, 1–12.
- Jastak, S., & Wilkinson, G. S. (1984). *Wide Range Achievement Test—Revised*. Wilmington, DE: Jastak Assessment Systems.
- Kaufman, A. S., & Kaufman, N. L. (1985). *Kaufman Test of Educational Achievement* (2nd ed.). Circle Pines, MN: AGS.
- *Kjeldsen, A.-C., Niemi, P., & Olofsson, A. (2003). Training phonological awareness in kindergarten level children: Consistency is more important than quantity. *Learning and Instruction*, 13, 349–365.
- *Kozminsky, L., & Kozminsky, E. (1995). The effects of early phonological awareness training on reading success. *Learning and Instruction*, 5, 187–201.
- *Kyle, F., Kujala, J., Richardson, U., Lyytinen, H., & Goswami, U. (2013). Assessing the effectiveness of two theoretically motivated computer-assisted reading interventions in the United Kingdom: GG rime and GG phoneme. *Reading Research Quarterly*, 48, 61–76.
- *Lie, A. (1991). Effects of a training program for stimulating skills in word analysis in first-grade children. *Reading Research Quarterly*, 26, 234–250.
- *Loeb, D. F., Gillam, R. B., Hoffman, L., Brandel, J., & Marquis, J. (2009). The effects of Fast ForWord language on the phonemic awareness and reading skills of school-age children with language impairments and poor reading skills. *American Journal of Speech-Language Pathology*, 18, 376–387.
- *Lyster, S. H. (2002). The effects of morphological versus phonological awareness training in kindergarten on reading development. *Reading and Writing: An Interdisciplinary Journal*, 15, 261–294.
- MacGinitie, W. H. (1978). *Gates–MacGinitie Reading Test: Canadian edition*. Toronto, Ontario, Canada: Nelson.
- *Mantzicopoulos, P., Morrison, D., Stone, E., & Setrakian, W. (1992). Use of SEARCH/TEACH tutoring approach with middle-class students at risk for reading failure. *Elementary School Journal*, 92, 573–586.
- Marulis, L. M., & Neuman, S. B. (2010). The effects of vocabulary intervention on young children’s word learning: A meta-analysis. *Review of Educational Research*, 80, 300–335. doi:10.3102/0034654310377087
- *Morris, R. D., Lovett, M. W., Wolf, M., Sevcik, R. A., Steinbach, K. A., Frijters, J. C., & Shapiro, M. B. (2012). Multiple-component remediation for developmental reading disabilities: IQ, socioeconomic status, and race as factors in remedial outcome. *Journal of Learning Disabilities*, 45, 99–127. doi:10.1177/0022219409355472
- *Nancollis, A., Lawrie, B., & Dodd, B. (2005). Phonological awareness intervention and the acquisition of literacy skills in children from deprived backgrounds. *Speech & Hearing Services in Schools*, 36, 325–335.
- Nation, K., & Coxsey, J. (2009). The relationship between knowing a word and reading it aloud in children’s word reading development. *Journal of Experimental Child Psychology*, 103, 296–308.
- National Early Literacy Panel. (Ed.). (2008). *Developing early literacy: A scientific synthesis of early literacy development and implications for intervention*. Denver, CO: National Institute for Literacy & Partnership for Reading.
- National Reading Panel. (Ed.). (2000). *Teaching children to read: An evidence-based assessment of the scientific research literature on reading and its implications for reading instruction*. Bethesda, MD: National Institute of Child Health and Development. Retrieved from <http://www.nichd.nih.gov/publications/nrp/report.htm>
- *O’Connor, R. E., Notari-Syverson, A., & Vadasy, P. (1998). First-grade effects of teacher-led phonological activities in kindergarten for children with mild disabilities: A follow-up study. *Learning Disabilities Research & Practice*, 13, 43–52.
- Paris, S. G. (2005). Reinterpreting the development of reading skills. *Reading Research Quarterly*, 40, 184–202.
- Pfost, M., Hattie, J., Dorfler, T., & Artelt, C. (2013). Individual differences in reading development: A review of 25 years of empirical research on Matthew effects in reading. *Review of Educational Research*. Advance online publication. doi:10.3102/0034654313509492
- *Phillips, L. M., Norris, S. P., & Mason, J. M. (1996). Longitudinal effects of early literacy concepts on reading achievement: A kindergarten intervention and five-year follow-up. *Journal of Literacy Research*, 28, 173–195.
- *Regtvoort, A. G. F.M., & van der Leij, A. (2007). Early intervention with children of dyslexic parents: Effects of computer-based reading instruction at home on literacy acquisition. *Learning and Individual Differences*, 17, 35–53.
- *Reitsma, P., & Wesseling, R. (1998). Effects of computer-assisted training of blending skills in kindergarteners. *Scientific Studies of Reading*, 2, 301–320.
- *Rothe, E., Grunling, C., Ligges, M., Fackelmann, J., & Blanz, B. (2004). Erste Auswirkungen eines Trainings der phonolo-

- gischen Bewusstheit bei zwei unterschiedlichen Altersgruppen im Kindergarten [First effects of training phonological awareness in two age cohorts of kindergarten children]. *Zeitschrift für Kinder- und Jugendpsychiatrie und Psychotherapie*, 32, 167–176.
- *Ryder, J. F., Tunmer, W. E., & Greaney, K. T. (2008). Explicit instruction in phonemic awareness and phonemically based decoding skills as an intervention strategy for struggling readers in whole language classrooms. *Reading and Writing: An Interdisciplinary Journal*, 21, 349–369.
- *Schachter, J., & Jo, B. (2005). Learning when school is not in session: A reading summer day-camp intervention to improve the achievement of exiting first-grade students who are economically disadvantaged. *Journal of Research in Reading*, 28, 158–169.
- Schmerkotte, H. (1978). Ergebnisse eines Vergleichs von Modellkindergärten und Vorklassen in Nordrhein-Westfalen [Results from a comparison of typical kindergartens and preschools in North Rhine-Westphalia]. *Bildung und Erziehung*, 31, 401–411.
- Scholin, S., & Burns, M. K. (2012). Relationship between pre-intervention data and post-intervention reading fluency and growth: A meta-analysis of assessment data for individual students. *Psychology in the Schools*, 49, 385–398.
- *Segers, E., & Verhoeven, L. (2005). Long-term effects of computer training of phonological awareness in kindergarten. *Journal of Computer Assisted Learning*, 21, 17–27.
- Seymour, P. H. K., Aro, M., Erskine, J. M., & COST Action Network. (2003). Foundation literacy acquisition in European orthographies. *British Journal of Psychology*, 94, 143–174.
- *Spörer, N., Brunstein, J. C., & Kieschke, U. (2009). Improving students' reading comprehension skills: Effects of strategy instruction and reciprocal teaching. *Learning and Instruction*, 19, 272–286.
- Stainthorp, R., & Hughes, D. (2004). What happens to precocious readers' performance by the age of eleven? *Journal of Research in Reading*, 27, 357–372.
- Stanford Achievement Test series: Technical data report* (8th ed.). (1990). San Diego, CA: Harcourt Brace Jovanovich.
- Stanovich, K. E. (1986). Matthew effects in reading: Some consequences of individual differences in the acquisition of literacy. *Reading Research Quarterly*, 21, 360–407.
- Suggate, S. P. (2009). School entry age and reading achievement in the 2006 Programme for International Student Assessment (PISA). *International Journal of Educational Research*, 48, 151–161.
- Suggate, S. P. (2010). Why “what” we teach depends on “when”: Grade and reading intervention modality moderate effect size. *Developmental Psychology*, 46, 1556–1579. doi:10.1037/a0020612
- Suggate, S. P. (in press). The parable of the sower and the long-term effects of early reading. *European Early Childhood Education Research Journal*.
- Suggate, S. P., Schaughency, E. A., & Reese, E. (2013). Children who learn to read later catch up to children who learn to read early. *Early Childhood Research Quarterly*, 23, 33–48. doi:10.1016/j.ecresq.2012.04.004
- Swanson, H. L., Hoskyn, M., & Lee, C. (1999). *Interventions for students with learning disabilities*. New York, NY: Guilford.
- Talbott, E., Wills, L., & Tankersley, M. (1994). Effects of reading comprehension interventions for students with learning disabilities. *Learning Disability Quarterly*, 17, 223–232.
- Therrien, W. J. (2004). Fluency and comprehension gains as a result of repeated reading: A meta-analysis. *Remedial and Special Education*, 25, 252–261.
- Thompson, G. B., McKay, M. F., Fletcher-Flinn, C. M., Connelly, V., Kaa, R. T., & Ewing, J. (2008). Do children who acquire word reading without explicit phonics employ compensatory learning? Issues of phonological recoding, lexical orthography, and fluency. *Reading and Writing: An Interdisciplinary Journal*, 21, 505–537.
- *Torgesen, J. K., Wagner, R. K., Rashotte, C. A., Herron, J., & Lindamood, P. (2010). Computer-assisted instruction to prevent early reading difficulties in students at risk for dyslexia: Outcomes from two instructional approaches. *Annals of Dyslexia*, 60, 40–56.
- Tunmer, W. E., Chapman, J. W., & Prochnow, J. E. (2004). Why the reading achievement gap in New Zealand won't go away: Evidence from the PIRLS 2001 International Study of Reading Achievement. *New Zealand Journal of Educational Studies*, 39, 127–145.
- Tunmer, W., & Greaney, K. (2009). Defining dyslexia. *Journal of Learning Disabilities*, 43, 229–243. doi:10.1177/0022219409345009
- *Vadasy, P. F., Jenkins, J. R., & Pool, K. (2000). Effects of tutoring in phonological and early reading skills on students at risk for reading disabilities. *Journal of Learning Disabilities*, 33, 579–590.
- *Vadasy, P. F., & Sanders, E. A. (2010). Efficacy of supplemental phonics-based instruction for low-skilled kindergartners in the context of language minority status and classroom phonics instruction. *Journal of Educational Psychology*, 102, 786–803. doi:10.1037/a0019639
- *Vadasy, P. F., & Sanders, E. A. (2011). Efficacy of supplemental phonics-based instruction for low-skilled first graders: How language minority status and pretest characteristics moderate treatment response. *Scientific Studies of Reading*, 15, 471–497.
- *Vadasy, P. F., & Sanders, E. A. (2012). Two-year follow-up of a kindergarten phonics intervention for English learners and native English speakers: Contextualizing treatment impacts by classroom literacy instruction. *Journal of Educational Psychology*, 104, 987–1005. doi:10.1037/a0028163
- *Vadasy, P. F., & Sanders, E. A. (2013). Two-year follow-up of a code-oriented intervention for lower-skilled first-graders: The influence of language status and word reading skills on third-grade literacy outcomes. *Reading and Writing*, 26, 821–843. doi:10.1007/s11145-012-9393-4
- *Vadasy, P. F., Sanders, E. A., & Peyton, J. A. (2006). Code-oriented instruction for kindergarten students at risk for reading difficulties: A randomized field trial with paraeducator implementers. *Journal of Educational Psychology*, 98, 508–528.
- *van der Kooy-Hofland, V. A. C., van der Kooy, J., Bus, A. G., van IJzendoorn, M. H., & Bonsel, G. J. (2012). Differential susceptibility to early literacy intervention in children with mild perinatal adversities: Short- and long-term effects of a

- randomized control trial. *Journal of Educational Psychology*, 104, 337–349. doi:10.1037/a0026984
- *van Keer, H. (2004). Fostering reading comprehension in fifth grade by explicit instruction in reading strategies and peer tutoring. *British Journal of Educational Psychology*, 74, 37–70.
- *Vaughn, S., Cirino, P. T., Linan-Thompson, S., Mathes, P. G., Carlson, C. D., Hagan, E. C., . . . & Francis, D. J. (2006). Effectiveness of a Spanish intervention and an English intervention for English-language learners at risk for reading problems. *American Educational Research Journal*, 43, 449–487.
- *Vaughn, S., Cirino, P. T., Tolar, T., Fletcher, J. M., Cardenas-Hagan, E., Carlson, C. D., & Francis, D. J. (2008). Long-term follow-up of Spanish and English interventions for first-grade English language learners at risk for reading problems. *Journal of Research on Educational Effectiveness*, 1, 179–214.
- *Vaughn, S., Linian-Thompson, S., Mathes, P. G., Cirino, P. T., Carlson, C. D., Pollard-Durodola, S. D., . . . & Francis, D. J. (2006). Effectiveness of Spanish intervention for first-grade English language learners at risk for reading difficulties. *Journal of Learning Disabilities*, 39, 56–73.
- *Vaughn, S., Mathes, P., Linian-Thompson, S., Cirino, P., Carlson, C., Pollard-Durodola, S. D., . . . & Francis, D. J. (2006). Effectiveness of an English intervention for first-grade English language learners at risk for reading problems. *Elementary School Journal*, 107, 153–180.
- Waddington, N. J. (1988). *Diagnostic reading and spelling tests: A book of tests and diagnostic procedures for children with learning difficulties*. Modbury, Australia: Waddington Educational Resources.
- Wagner, R. K., Torgesen, J. K., & Rashotte, C. A. (1999). *Comprehensive Test of Phonological Processing*. Austin, TX: PRO-ED.
- Wanzek, J., Vaughn, S., Kim, A., & Cavanaugh, C. L. (2006). The effects of reading interventions on social outcomes for elementary students with reading difficulties. *Reading & Writing Quarterly*, 22, 121–138.
- *Warrick, N., Rubin, H., & Rowe-Walsh, S. (1993). Phoneme awareness in language-delayed children: Comparative studies and intervention. *Annals of Dyslexia*, 43, 153–173.
- Wiederholt, J. L., & Bryant, B. R. (1992). *Gray Oral Reading Test-3*. Austin, TX: PRO-ED.
- *Williams, J. P. (1980). Teaching decoding with an emphasis on phoneme analysis and phoneme blending. *Journal of Educational Psychology*, 72, 1–15.
- *Wolff, U. (2011). Effects of a randomised reading intervention study: An application of structural equation modelling. *Dyslexia*, 17, 295–311. doi:10.1002/dys.438
- Woodcock, R. W. (1998). *Woodcock Reading Mastery Tests—Revised normative update Forms G and H examiners manual*. Minneapolis, MN: Pearson Assessments.