HHS Public Access

Author manuscript

*Rev Educ Res*. Author manuscript; available in PMC 2015 November 01.

Published in final edited form as:

*Rev Educ Res*. 2015 September 1; 85(3): 395–429. doi:10.3102/0034654314555996.

# Are Child Cognitive Characteristics Strong Predictors of Responses to Intervention? A Meta-Analysis

**KARLA K. STUEBING, PhD [research professor]**

Texas Institute for Measurement, Evaluation, and Statistics, Department of Psychology, University of Houston, 4811 Calhoun Rd, 4th Floor, Houston, TX 77204

**AMY E. BARTH, PhD [assistant professor]**

Department of Special Education at the University of Missouri-Columbia, 311B Townsent Hall, Columbia, MO 65211

**LISA H. TRAHAN, MA [graduate student]**

Department of Psychology at the University of Houston, 4811 Calhoun Rd, Houston, TX 77204

**RADHIKA R. REDDY, MA [graduate student]**

Department of Psychology at the University of Houston, 4811 Calhoun Rd, Houston, TX 77204

**JEREMY MICIAK, PhD [research assistant professor]**

Texas Institute for Measurement, Evaluation, and Statistics, Department of Psychology, University of Houston, 4811 Calhoun Rd, Houston, TX 77204

**JACK M. FLETCHER, PhD**

Hugh Roy and Lillie Cranz Cullen Distinguished Professor of psychology at the University of Houston, 4811 Calhoun Rd, Houston, TX 77204

## Abstract

We conducted a meta-analysis of 28 studies comprising 39 samples to ask the question, "What is the magnitude of the association between various baseline child cognitive characteristics and response to reading intervention?" Studies were located via literature searches, contact with researchers in the field, and review of references from the National Reading Panel Report. Eligible participant populations included at-risk elementary school children enrolled in the third grade or below. Effects were analyzed using a shifting unit of analysis approach within three statistical models: cognitive characteristics predicting growth curve slope (Model 1, mean $r = .31$), gain (Model 2, mean $r = .21$), or postintervention reading controlling for preintervention reading (Model 3, mean $r = .15$). Effects were homogeneous within each model when effects were aggregated within study. The small size of the effects calls into question the practical significance and utility of using cognitive characteristics for prediction of response when baseline reading is available.

karla.stuebing@times.uh.edu. bartha@missouri.edu. isahtrahan@ gmail.com. radhika7star@ gmail.com. Jeremy.Miciak@times.uh.edu. Jack.Fletcher@times.uh.edu.

**Keywords**

response to intervention; reading disabilities; reading instruction; meta-analysis

Over the past three decades, there has been a significant and expanding interest in understanding factors that underlie the development of reading skills and instructional response in the context of reading intervention. This research was built on earlier studies that helped establish learning to read as a complex set of proficiencies supported by different cognitive skills that vary depending on the domain of reading that is addressed. To illustrate, the importance of phonological awareness for the development of word recognition skills (Bryant, MacLean, Bradley, & Crossland, 1990; Byrne, Fielding-Barnsley, Ashley, & Larsen, 1997; Speece, Ritchey, Cooper, Roth, & Schatschneider, 2004) and of world and word knowledge for reading comprehension are well-established (Carroll, 1993; Catts, Adlof, & Weismer, 2005; Kendeou, Savage, & van den Broek, 2009; Vellutino, Tunmer, Jaccard, & Chen, 2007). Other cognitive skills such as rapid naming and working memory are also implicated as factors in learning to read (Vellutino et al., 2007).

Less studied to date is to what extent these cognitive skills are related to growth in reading skills either through development or through treatment response in reading intervention studies. In the former, the population is usually unselected and often occurs in a longitudinal context in which cognitive tasks are used to predict which students may struggle with learning to read (Schatschneider, Fletcher, Francis, Carlson, & Foorman, 2004). In the latter, the population typically studied comprises children identified as at risk for reading failure in the context of ongoing reading instruction. Often, these studies compare the relative effects across predictors for children who respond adequately and inadequately to instruction to help identify unique characteristics of inadequate responders.

Understanding the role of specific cognitive skills and estimating the magnitude of their effects in both of these assessments of growth may help improve prediction of risk status and individual response to instruction and could facilitate the development of interventions tailored to individual differences across learners. Such efforts take on special importance as reading intervention research focuses more intently on students who do not respond adequately to instruction, especially in the context of response to intervention service delivery frameworks that depend on both the detection of risk and inadequate response (Fletcher & Vaughn, 2009).

Despite these potential contributions to reading assessment and instruction, there is controversy and uncertainty surrounding the nature and utility of different cognitive skills as predictors of reading growth or intervention response. In studies of reading intervention, there is disagreement on the unique contribution of cognitive skills and the utility of cognitive assessment, especially in relation to predicting or treating inadequate responders. Some argue that assessment of cognitive skills is critical in children who have not responded to intervention because such assessments permit instruction tailored to individual student needs (Decker, Hale, & Flanagan, 2013; Hale, Fiorello, & Thompson, 2010). For example, Decker et al. (2013) argued that the inclusion of cognitive assessments facilitates understanding of individual learning differences that directly affect the efficacy of academic interventions.

In contrast, others observe the general lack of evidence that patterns of strengths and weaknesses in cognitive attributes interact with treatment outcomes (Fuchs, Hale, & Kearns, 2011; Kearns & Fuchs, 2013; Pashler, McDaniel, Rohrer, & Bjork, 2009). Kearns and Fuchs (2013) reported that although cognitively focused instruction is more effective than Tier I delivered by general education teachers, cognitively focused instruction was not more effective than rigorous academically focused instruction. Less controversial is the evidence that specific cognitive skills can serve as precursors and predict those at risk for reading failure (Scarborough, 1998) although there is disagreement and uncertainty about what variables are most predictive. The available evidence suggests that four cognitive constructs best predict at-risk status as well as intervention response (Fletcher, Lyon, Fuchs, & Barnes, 2007): phonological awareness, isolated or rapid letter naming, verbal working memory, and oral language/vocabulary. There is also emerging evidence that these same four skills best differentiate adequate and inadequate responders to instruction (Denton et al., 2013; Fletcher et al., 2011; Vellutino, Scanlon, Small, & Fanuele, 2006). However, there are conflicting data on which of these correlated language-based measures are the strongest predictors and under what circumstances inclusion of them in an assessment battery might be useful. These are important questions for designing parsimonious assessments and understanding underlying relations.

In part, these conflicting data result from different approaches to modeling growth and estimating the magnitude of individual variable contributions to reading outcomes, historically a significant methodological issue in multiple domains of behavioral research (Darlington, 1968). In the section that follows, we review recent summaries of research pertinent to these questions, highlighting differences in methodologies that may explain differences in findings. We then highlight lingering questions addressed in the present meta-analysis.

## Estimates of Relations Between Cognitive Predictors and Reading Outcomes

Several summaries of research have addressed the relation between cognitive predictors and reading outcomes, as well as the magnitude of those relations within different study designs. These are not all the relevant studies and were selected to demonstrate differences in how change is modeled. For example, Scarborough (1998) summarized literature predicting reading outcomes from baseline learner characteristics (BLCs) measured before or close to the beginning of reading instruction with a criterion of reading measured after 1 to 3 years of instruction. The most robust correlations were moderate, with mean bivariate correlations for isolated letter identification ($M$ = 0.52), phonological awareness ($M$ = 0.46), Full-Scale IQ ($M$ = 0.41), and rapid naming ($M$ = 0.38), among others. The samples included in these analyses were unselected, representing the full range of achievement on both predictor and criterion variables. Thus, because there is minimal restriction of range the correlations obtained should be larger than those that would be found within intervention studies where students are initially selected for risk of reading failure. However, the longtime gap between assessment of the BLC and the reading outcome will most likely reduce the observed correlation.

Swanson, Trainin, Necoechea, and Hammill (2003) performed a meta-analysis (Hunter & Schmidt, 1990) of the relation of phonological awareness (PA) and rapid naming (RAN) with word reading in a test of the double deficit hypothesis of reading disability (Wolf & Bowers, 1999). They aggregated correlations between RAN and PA with word reading outcomes, correcting the observed correlations for unreliability, restriction of range, and sampling error. They selected only studies where the relevant assessment of these variables was done within a 1-month time window. The meta-analysis included correlations for low performing groups, high performing groups, and mixed groups. The average correlations of PA and RAN with reading were moderate ($M$ = 0.48 and $M$ = 0.46, respectively) and were lower in the lower performing groups even after correcting for restriction of range ($M$ = 0.30 and $M$ = 0.41 for PA and RAN, respectively, for low performing groups, and $M$ = 0.56 and $M$ = 0.43 for skilled/average readers).

Nelson, Benner, and Gonzalez (2003) estimated the "strength and relative magnitude of the influence of the learner characteristics on the treatment effectiveness of early literacy interventions" (p. 256). They began with a group of 22 studies reviewed by Al Otaiba and Fuchs (2002) and added 11 additional studies for a meta-analysis. We assume that treatment effectiveness was operationalized as change in reading performance and we would expect that the correlations included in this meta-analysis should be between BLCs and reading growth parameters or gain scores although this is not explicitly stated in the article. The analysis included studies of students at risk for reading disabilities due to initial low ability, low PA, low income, other disabilities, or language disorders. Because the sample was selected and therefore demonstrates an uncorrected restriction of range on both predictors and criteria, lower effect sizes than Scarborough (1998) and Swanson et al. (2003) would be expected. Nelson et al. (2003) reported mean weighted Fisher's $z$, which when converted to correlations with reading outcomes are .40 for PA, .25 for Full-Scale IQ, and .47 for RAN. These findings are somewhat consistent with what would be expected given range restriction and are consistent with results reported by Swanson et al. (2003), but it is not clear that comparing them is appropriate because they ask different questions.

What is important to note about the mean correlations reported in these three studies is that they do not represent the same parameter. The correlations in Scarborough (1998) are between a BLC at Time 1 and a reading outcome at Time 2, 1 to 3 years later. In Swanson et al. (2003), the correlations are for measures taken within 1 month of each other. We assume that in the Nelson et al. (2003) meta-analysis, given their description, the correlations are between the BLC at Time 1 and the amount of growth that occurred during the intervention period. These studies answer three different questions, and it is important that we select the effect size from the correct set of studies when using them to develop interventions or assessments. We should not assume that a predictor that shows a moderate correlation with final status at one time point is predictive of reading level at a later time point or that it predicts growth in general or growth in response to intervention.

As an example of a study where the effects of a BLC were analyzed separately by analytic model, Stuebing, Barth, Molfese, Weiss, and Fletcher (2009) completed a meta-analysis of 22 studies that examined the predictive power of preintervention IQ scores. The mean *r* between IQ and reading outcome was .27. In models where only the pretest score and IQ were included as predictors, IQ uniquely accounted for approximately 3% of the variance in reading outcomes, which is comparable to a semipartial correlation of .17. This meta-analytic estimate is much lower than in Scarborough (1998) and Nelson et al. (2003), but is it because of the particular BLC studied or because a different parameter was being estimated? In models where BLCs other than IQ were also included as predictors, IQ accounted for about 1% (semipartial $r = .1$) of the unique variance in growth during the intervention, indicating that IQ was not a robust predictor of intervention response. This finding is consistent with other evidence showing that IQ is a weak predictor of long-term growth in reading ability (Share, McGee, & Silva, 1989) and meta-analytic evidence that IQ-discrepancy is not a reliable marker of specific reading disability (Stuebing, Fletcher, LeDoux, & Lyon, 2002).

## Summary and Lingering Questions

These summaries show that PA and RAN are consistently identified as very good predictors of reading outcomes based on bivariate correlations in selected and unselected populations, accounting for approximately 20% to 25% of the variance in reading in prediction and intervention studies. These correlations compare favorably to the predictive power of IQ, which, based on a correlation of .27, accounts for about 7% of the variance in response to intervention in reading. However, whether PA or RAN is the better predictor remains unclear both in terms of the magnitude of association with response to intervention as well as in terms of whether either of them have unique predictive power relative to each other and to other potential predictors. It may be that they are roughly comparable and that the differences across studies reflect sampling variation or measurement issues. What is not known is whether these BLCs remain strongly associated with response to intervention when reading level at pretest is controlled and whether the relations hold across different methods for modeling change.

These differences across studies have led to controversies over the nature of reading development and difficulty. For example, is the contribution of PA sufficient in itself to explain variation in reading development and reading difficulties, or does the emergence of RAN indicate that a theory based on phonological processing is inadequate and must incorporate theory on what skills are measured by rapid naming tests (Wolf & Bowers, 1999). This difficulty led to the double deficit hypothesis of reading disability, which suggests that PA and RAN represent independent and separable contributors to reading difficulties (Wolf et al., 2003). RAN, however, is most predictive for alphanumeric symbols, which may indicate that it is a measure of early reading prominence and related to the evidence that knowledge of letter names and sounds is a strong kindergarten predictor of reading outcomes (Scarborough, 1998; Schatschneider et al., 2004). For IQ, estimates of the magnitude vary widely across studies, leading some to suggest that IQ is a robust predictor (Naglieri, 2001) and others to suggest that the contribution is weak (Vellutino, Scanlon, &

Lyon, 2000). Such conflicting conclusions fuel debate regarding the utility of IQ assessment among reading researchers and in special education communities.

However, much of the inconsistency found in summaries of this literature may be attributable to methodological issues in study design and variations in the model of change employed at the study or meta-analytic level. In order to make informed decisions about the use of BLCs, including cognitive skills, as predictors of reading outcomes for students in the general population and in intervention, it is important to accurately estimate the magnitude of their effects across studies. To accomplish this goal, only effects that represent the same underlying parameter from comparable models should be combined across studies for interpretation. As we shall see, the magnitude of effects in previous summaries of research is often estimated from models that are not pertinent for questions about the predictors of response to intervention.

## Models of Change

Different combinations of predictors (e.g., BLC alone, with other BLCs or with a reading pretest), outcomes (e.g., posttest status, slopes from growth curve models, or responsiveness designation), time-sampling, and population selection allow for estimation of different parameters and effects to represent the relations between a BLC and reading outcome. These combinations can be conceptualized as variations in the underlying model of change. In the sections that follow, three distinct models of change are specified, which subsequently guide the present meta-analysis. Studies falling within one of these models of change will result in like effects that can be aggregated. These models include two unconditional models and one conditional model. Importantly, the identification of distinct models of change is not evaluative. Rather, the models address different research questions. Because change is most explicitly operationalized in the context of intervention, we focus on this class of studies. The bivariate models underlying the studies summarized by Scarborough (1998) and Swanson et al. (2003) are not conceptualized as studies of change and thus are not developed further in this meta-analysis but are a useful context for our results.

### Model 1: Unconditional Growth Curve Models

Growth in reading in response to intervention via growth curve models can be assessed if reading skills are assessed over three or more time points. Within studies using growth curve models, various BLCs are used to account for variance in the latent slope or growth in a reading skill over the course of the intervention. In the present meta-analysis, the effect size collected from growth curve studies was a bivariate correlation between the estimated slope and the BLC. To the extent that either measure has less than perfect reliability, the estimate of the correlation will be attenuated. A drawback of this model, as with any correlational model, is that the source of the observed covariation is left unexplained. To the extent that there are other causes of growth that are correlated with the BLC but omitted from the model, the resulting correlation will be biased, potentially becoming spuriously high. As in any case of model misspecification, this bias is not evaluable with statistical methods but depends on the validity of the underlying theoretical model of causes and effects.

### Model 2: Unconditional Gain Models

A second type of model frequently used to examine improvements in reading development over time is a gain model (Model 2) in which subjects are measured at only two time points. With a gain model, BLCs are used to predict observed gain in reading performance over time (pretest performance is subtracted from posttest performance). Similar to growth curve models, the critical issue is quantifying the amount of change in reading skill over the course of intervention and isolating what predicts it. Although gain is an unbiased estimator of growth in reading skill, it is less reliable than growth curves based on three or more data points (Rogosa, Brandt, & Zimowski, 1982). As a consequence, the true effects from Model 2 studies should have the same approximate magnitude of relations as with growth curve models, but the observed relations will be more attenuated due to greater unreliability in the gain indicator. The same potential for model misspecification and biased effect sizes exists for both Models 1 and 2. Both are models of the unconditional correlates of change

### Model 3: Change Conditioned on Initial Status

In this common model, the posttest is predicted by both the pretest and a BLC. This model differs from the first two models in two ways. First, the outcome is postintervention status rather than gain or growth. If this were the only difference and if raw scores or interval (e.g., Rasch) scores rather than age-adjusted standard scores were used for pretest and posttest, we might expect effects of similar magnitudes to those found for Models 1 and 2. Especially in situations where the initial status is highly restricted (as is the case for many samples where subjects are selected due to low performance at the initial point of measurement), the gain/growth variable tends to be quite highly correlated with posttest.

However, the model also differs because an additional predictor (i.e., pretest) is included, which changes the nature of the relation from unconditional to conditional and changes interpretation to that of a relation among residuals. The numerator of the semipartial correlation is the covariance between the residual of posttest controlling for pretest and the residual of BLC controlling for pretest. This covariance is then standardized through division by the product of the standard deviation of the posttest and the standard deviation of the residuals of the BLC. Importantly, the numerator of this index is a covariance of two residuals, or the parts of both posttest and the BLC that are not linearly related to pretest. This relation is typically described as that which exists when pretest is controlled. A positive correlation would imply that an individual who is higher on the BLC than would be expected based on performance on the pretest would also be higher than expected on the posttest even after taking pretest into account. Because a semipartial correlation does not estimate the same parameter as the bivariate correlation, analyses should be completed separately for both conditional and unconditional model classes.

### Commonalities Among Models of Change

Despite important analytical differences, the two unconditional models and the conditional model all represent change models. All three models use the BLC to predict a measured gap. This is most easily understood in Model 2, in which the gap between pretest and posttest is predicted from the BLC. In Model 1, the gap is the average amount of difference between adjacent time points (slope), which is correlated with the BLC. In Model 3, the gap is the difference between the predicted level of posttest (given pretest) and the actual posttest. This difference is then correlated with the residualized BLC.

It is critical that the two research questions underlying both conditional and unconditional models be well understood. Hand (1994) encouraged deconstructing the statistical question and relevant latent assumptions to assure that they parallel the question the researcher wishes to answer. Models 1 and 2 (the unconditional models) ask the question, "Is a linear relation observed between growth in response to intervention and some BLC?" This relation is not conditioned on any other variables, including initial status. A high value would indicate that students high on the BLC across the spectrum of pretest scores would be expected to have high growth of approximately the same magnitude. The nature of the relation is not unpacked; there is no necessary causal relationship specified and spuriousness is not ruled out. Model 3 asks a different question. A researcher who observes a high correlation from Models 1 or 2 might wonder if the relation between growth and BLC is partially explained by initial status. Model 3 directly addresses this question and also allows us to determine if the BLC provides information that is not already contained in initial status.

Across research questions, there are two common purposes for research investigating correlates of intervention response. First, such research seeks and refines explanatory factors to elaborate a theoretical model of how children learn to read, which is then used to build the intervention. The second purpose is to establish accurate predictive equations that can be used in educational settings to predict risk, inform instructional decisions, and perhaps improve treatment. In the first case, matters of spurious relations are important for theoretical clarity and should be sorted out. The finding of an initial bivariate relation should almost certainly be followed with studies that investigate the causal direction of the effects and also test for additional variables that might explain the effects, either as mediators or as causes of both growth and the BLCs. However, if the purpose of the study is to find additional variables that improve prediction of final status in the presence of an intervention, the theoretical model of relations among variables is less important except as a source of plausible candidate predictors.

## Purpose

The purpose of this study was to examine the extent to which the cognitive processes underlying word recognition and reading comprehension predict growth in reading, operationalized here as intervention response. Such processes have been demonstrated to be highly predictive of intervention response using postintervention status as an outcome. However, less is known about how such processes predict growth in reading, both in an unconditional sense as represented by bivariate correlations between BLCs and gain/growth and in a conditional sense as represented by the semipartial correlations of a BLC with posttest controlling for initial status. This meta-analysis focused on three main questions:

1. What is the magnitude of the correlations between BLCs and growth in word reading, reading fluency, and reading comprehension over the course of intervention among early grade readers (Model 1 and Model 2)?

2. To what extent can BLCs be used to predict the gap between actual postintervention status and postintervention status predictable from initial status alone (Model 3)?
3. What is the relative predictive power of different cognitive measures?

We predicted that within a BLC, the largest correlations should be found in the unconditional growth models (Models 1 and 2). We predicted that the growth curve models (Model 1) will have the largest estimated effects because of the higher reliability of the estimated effects. Gain models (Model 2) will have smaller estimated effects due to the comparatively lower reliability of the gain score. The conditional growth model (Model 3) will likely have the smallest correlations, both because of unreliability in the three measures involved, and because the effect is a semipartial correlation rather than a bivariate correlation. Under normal patterns of bivariate correlations among the three measures, the semipartial correlation will be smaller than the bivariate correlation. We also predicted that measures of PA would be slightly more predictive of intervention response than measures of RAN, and that both would be stronger than measures of verbal short-term memory and vocabulary. None of the cognitive measures was expected to be more predictive than measures with a print component, such as letter recognition, spelling, and reading (Vellutino et al., 2006).

## Method

### Search Strategies

Several methods were used to gather studies pertaining to response to reading intervention and published in peer-reviewed journals, including literature searches, contact with researchers in the field, and review of references from the National Reading Panel (NRP) Report. The search terms “reading,” “response to reading intervention,” “response to treatment intervention reading,” and “reading interventions” were entered into Google Scholar. A literature search was also completed on PsycINFO using the search terms “reading” and “response to intervention” or “response to instruction.” In cases where articles were deemed relevant but included insufficient information to calculate an effect size, an email inquiry was sent to the first author to request use of the data. Four authors provided data for five separate studies, and models were constructed directly from the data. This effort permitted the computation of effects not reported in published studies. The search time frame was from 1995 to 2010, inclusive.

The NRP (National Institute of Child Health and Human Development, 2000) published a report investigating the effectiveness of early reading intervention. A search was conducted to gather articles by authors whose work was included in the meta-analysis described in the special report. Searches were also conducted for authors whose studies were ultimately excluded from the NRP meta-analysis or whose work appeared in the reference list. These searches included the author’s last name and the search term “reading.”

Two graduate students in psychology reviewed abstracts for every article whose title hinted at a reading intervention. If the abstract confirmed that the study pertained to a reading intervention, the full text was reviewed by both graduate student coders in order to determine whether the children met age- and language-related criteria, and if so, whether sufficient information was reported to calculate an effect size. If at least one coder found that the study should be included, it moved ahead in the process.

## Inclusion and Exclusion Criteria

The primary research questions pertain to the extent to which individual characteristics at baseline can predict growth in reading during the intervention period. Therefore, outcome variables assessing response to intervention included measures of fluency, reading comprehension, and word reading, which captured posttest performance, performance gain, or growth in performance. Potential BLCs included scores on oral language, phonological awareness, rapid naming, spelling/orthographic processing, and working memory at baseline.

Eligible participant populations included at-risk elementary school children enrolled in the third grade or below. To avoid potential confounds related to language impairments or second language acquisition, samples of children who were severely language impaired were excluded. Furthermore, participants had to be proficient in English; thus, studies comprising students who were primarily or exclusively English language learners or in English as a second language programs were excluded. Studies were also excluded if the sample was exceptional in any way; for example, some samples were composed entirely of children with speech articulation difficulties or disorders or behavior problems. To the extent that the authors were able to determine that the intervention represented a second or higher attempt at remediation, the study was excluded.

## Coding

Any study reporting results of an intervention designed to improve reading-related skills in at-risk populations of children was considered for eligibility. A study was eligible if the results provided enough information to estimate an effect size for at least one of the three relevant models. This set of models was derived inductively as coding was performed: correlates of growth curve (Model 1), correlates of gain (Model 2), and conditional correlates of gain (Model 3).

Effect sizes were coded in different ways and with different degrees of estimation depending on the data provided in each study. We were able to code effects from raw data (occasionally reported in studies with small *N*s or provided by the authors), from correlation matrices, and from test statistics. Occasionally, a text described a pertinent result as being significant or nonsignificant, which permitted estimation of the upper and/or lower bound of the effect size. A lower bound estimate for a nonsignificant result was estimated at 0 and the upper estimate was the largest nonsignificant $r$ that might be obtained given the degrees of freedom for that model. Results reported as significant but for which no quantitative data were provided were estimated as the smallest possible significant correlation given the degrees of freedom in the study. No upper bound was estimated in these cases. Each effect size was coded for the degree of estimation required to calculate it. For example, provision of data by the authors of a study required no estimation in constructing the models and calculating the effect sizes. However, results reporting the significance of the effect without the provision of quantitative data to support the statement required calculating a range for the effect size and thus represented a high degree of estimation. Degree of estimation was assessed on a Likert-type scale ranging from 1 (*High estimation*) to 5 (*No estimation*). Examples for each degree of estimation are illustrated in Table 1. Estimation was included as a moderator variable in each analysis so that we could assess the sensitivity of our results to coding assumptions.

Coding was completed by two graduate students in psychology who had completed graduate-level courses in statistics and training with the first author, including review of relevant texts and practice coding. Each student attempted to code all articles judged to be tentatively eligible. At meetings attended by both students and the first author, the students compared the models coded and the effect sizes they calculated for each model. Agreement was calculated for each study and determined by the number of effect size values agreed on. In cases where the effect size values differed, consultation with the first author resulted in selection of the appropriate effect size value. In cases where models differed, the students coded the extra models and met again to compare the effect size values.

Statistical moderators included type of model, degree of estimation and the specific BLC used to predict response. Two authors highly familiar with cognitive constructs coded the particular measures used as predictors in each study into a smaller set of constructs to allow for aggregation of results. For example, both "reading of real words" and "reading of nonsense words" were coded under the Word Reading construct. "Rapid automatized naming (RAN) of letters" and "rapid automatized naming of colors" were coded under the RAN construct. See the appendix for a list of measures for each construct.

## Analysis

Comprehensive Meta-Analysis software (Borenstein, Hedges, Higgins, & Rothstein, 2005) was used for the core set of analyses, using the module that requires input of an effect size and its variance for each study. Correlation coefficient effect sizes converted to Fisher's $z$ ($Z_r$) were chosen to represent the strength of association between BLCs and outcomes representing response to intervention. The variance of each $Z_r$ was estimated by the following formula from Borenstein, Hedges, Higgins, and Rothstein (2009): $\mathrm{Var}_{Zr} = 1/(N - 3)$. We chose to analyze these effects using a random effects model rather than a fixed effects model because it did not seem plausible that these effects would arise from a distribution with one true common effect where all of the variability was due to sampling error. The studies varied in many ways including the analytic approach used, the population from which the sample was drawn, the type, duration, and intensity of intervention delivered, and the particular combination of baseline characteristic and outcome variable studied. Due to these and many other influences, it was more plausible that the variance in the effects we observed represented both random variance about some true set of means and also true variance of the means due to a myriad subtle influences.

A shifting unit of analysis approach was used to handle the dependencies in the effects from the same study (Cooper, 1998). Within models, effects were analyzed in two ways ignoring BLC. The first way included all effects from each study, and the second included an average effect from each study resulting in independent effects. Finally, independent effects were analyzed within model and within BLC. The number of studies within model for the unaggregated analysis and the analysis taking BLC into account is the same and thus provides transparency.

A sensitivity analysis was carried out using meta-regression on all effects from each study (unaggregated) to ascertain whether the estimation process we used in coding effects biased our results. Meta-regression is typically used to account for significant heterogeneity in effects with a continuous moderator variable, but in this case, we used it to be sure that effects with high levels of estimation were not systematically different from those that were coded directly from the studies. A significant slope of effect size on degree of estimation could indicate bias. We also plotted the results of these analyses to allow for inspection of potential nonlinear relationships between degree of estimation and effect size.

## Results

Screening of titles and abstracts resulted in 129 candidate studies. From these, we were able to obtain 120 effects from 28 studies that included 39 separate samples. All studies were independently coded by the third and fourth authors. Initial agreement was .91 when agreement within studies was averaged and it was .96 when agreement over all data points was calculated. All disagreements were resolved by the team. Mean meta-analytic effects by model with results aggregated and unaggregated within study are presented in Table 2, but only the unaggregated results are presented below, because with the exception of minor difference in the width of the confidence intervals and one difference in the test of heterogeneity, the results are redundant. The results by BLC within model are presented in Table 3.

### Model 1

The test of homogeneity for the 36 effects using growth curve analyses (Model 1) was significant, $Q_{(35)} = 51.84$, $p = .03$, which lent some support to our choice of a random effects model. The average effect in a random effects model aggregating over all BLCs correlated with response to instruction as growth curves was $r = .31$, 95% confidence interval (CI) = .27 to .36 (see Table 2). We evaluated the effect of estimation on our conclusion about these effects by carrying out a meta-regression moderator analysis with effect size predicted by level of estimation. Meta-regression examines whether particular moderator variables explain the heterogeneity of study effects. The effect of estimation was not significant in an unrestricted maximum likelihood mixed effects meta-regression, $b = .004$, $Q_{(1)} = 0.074$, $p = .79$, indicating that the magnitude of the effects was not linearly related to amount of estimation.

Figure 1, a visual display of the meta-regression relation between the effects and level of estimation, shows that the mean effect at each level of estimation is approximately the same. There is no systematic difference between effects coded with minimal input from the article (high estimation seen on the left hand side of the display) and those coded with complete data (less estimation on the right hand side of the display).

A fixed effects analysis using BLC as a categorical moderator variable was significant, $Q_{(9)}$ = 19.64, $p$ = .02, and the residual, within BLC variance was not significant, $Q_{(26)}$ = 32.19, $p$ = .19, suggesting that within Model 1 with unaggregated effects, the significant heterogeneity is explained by BLC. The mean effects by BLC are reported in Table 3. These results are independent within BLC, with each effect coming from a separate group of subjects so the confidence intervals should be correct. Growth within the intervention period was related to Phonological Awareness ($r$ = .31, 95% CI = .26-.37), RAN ($r$ = .34, 95% CI = .24-.43), Word Reading ($r$ = .35, 95% CI = .24-.45), Oral Language ($r$ = .19, 95% CI = .09-.28), and Spelling/Orthography ($r$ = .38, 95% CI = .24-.45). Summarily, over multiple studies, these BLCs predicted 4% to 14% of the variance in growth. There was little difference in the predictive utility of the different cognitive constructs except for the lower relation of measures of oral language.

We followed up the significant effect of BLC with an exploration of the source of the heterogeneity. Because the confidence intervals around the mean effect for each BLC except oral language included the overall mean effect, we reran the analysis within this model, excluding the five effects where oral language was the BLC. In this case, the test using BLC as a categorical moderator was no longer significant, $Q_{(8)}$ = 13.55, $p$ = .09. The test of the residual within BLC variance remained nonsignificant, $Q_{(22)}$ = 26.34, $p$ = .24. The mean effect over this set of studies was $r$ = .33, 95% CI = .29-.37. In the Model 1 analysis using the independent effects aggregated within studies, the heterogeneity was not significant $Q_{(13)}$ = 14.20, $p$ = .36, so no follow-up was pursued.

## Model 2

The test of homogeneity for the 30 unaggregated effects correlating pretest–posttest gain with BLCs (Model 2) was not significant, $Q_{(29)}$ = 14.18, $p$ = .99, failing to provide support for use of a random effects model. The estimate of true variance was $I^2$ = .00. Estimated $\tau$ was 0, suggesting that all of the observed variance in this set of effects is consistent with sampling error. The average fixed effect over all BLCs with response to instruction operationalized as gain was $r$ = .22, 95% CI = .16-.27. Again, there were dependencies within this set of effects, so the confidence interval should be interpreted with caution because it is likely too narrow. The confidence intervals for the aggregated effects are accurate.

The sensitivity analysis with effect size predicted by estimation was not significant, $b$ = −.007, $Q_{(1)}$ = 0.08, $p$ = .77, indicating that the magnitude of the effects was not significantly linearly related to the degree of estimation (Figure 2). The regression line has a nonsignificant negative slope, indicating that descriptively, the effects with small amounts of estimation (right hand side of the display) are a bit smaller than those that required more estimation (left hand side of the display.) However, because of the nonsignificance of the slope, the lack of heterogeneity in the data, and the fact that there are very few effects that required high levels of estimation, we are not concerned that the overall effect size is unduly upwardly biased by our estimation.

Despite the lack of heterogeneity, we descriptively presented the results of each BLC separately (Table 3). These results are independent within BLC with each effect coming from a separate group of subjects so the confidence intervals should be correct. Despite the apparent differences in mean effects across BLCs for Model 2, it is most appropriate given these results to treat them as homogeneous, that is, the best estimated correlation for each of them is the meta-analytic mean. Thus, there is no apparent difference in the predictive power of the different constructs. Across multiple studies, these BLCs accounted for 2.6% to 6% of the variance in gain. The relations in general were weaker than for growth curves, possibly because gain is measured less reliably than growth.

### Model 3

The test of homogeneity for the 54 effects predicting status on posttest from status on pretest and BLC (Model 3) was not significant $Q_{(53)} = 35.98$, $p < .97$, again failing to provide support for the use of a random effects model. The average fixed effect over all BLCs for the partial correlation between BLC and posttest, conditional on pretest was $r = .15$, 95% CI = .11 to .19. Dependencies in effects within this set of studies will tend to shrink the CIs so they should be interpreted with caution. The confidence intervals for the aggregated effects are accurate. The meta-regression (sensitivity analysis) with effect size predicted by estimation was not significant, $b = -.02$, $Q_{(1)} = 0.71$, $p = .40$, indicating that the magnitude of the effects was not a result of bias due to linear effects of estimation (Figure 3). The slope is slightly negative again but nonsignificant. What is more, the distributions of effects within the highest and lowest levels of estimation (far left and far right in the display) appear to be very similar, leading us to conclude that there is little bias due to estimation in the mean effect. Both $I^2$ and $\tau$ were estimated at 0 for this model indicating that the observed heterogeneity is consistent with sampling error and not true variance.

The average semipartial correlations of each BLC with posttest are presented (Table 3). These results are independent within BLC with each effect coming from a separate group of participants, so the confidence intervals should be correct. After controlling for pretest, posttest was related to Phonological Awareness ($r = .14$, 95% CI = .09-20), RAN ($r = .30$, 95% CI = .21-.38), Word Reading ($r = .11$, 95% CI = .05-.17), Oral Language ($r = .14$, 95% CI = .07-.21), Spelling/Orthography ($r = .18$, 95% CI = .09-.26), and Nonverbal ($r = .11$, 95% CI = −.04 to −.26). Despite the apparent descriptive differences in mean effects across BLCs, the most defensible use of these data is to use the overall mean estimate for each BLC because the estimate of the true variance or $\tau$ was 0. Across multiple studies, these BLCs uniquely account for 1% to 9% of the variance in change controlling for pretest, with the average variance accounted for equal to 2.25%. The relations in general are weaker than those we found for the unconditional bivariate correlations with growth curves or gain.

## Discussion

Three previous reviews of the literature using meta-analytic techniques examined the extent to which broad cognitive characteristics of the learner predict reading level or response to reading intervention (Nelson et al., 2003; Scarborough, 1998; Swanson et al., 2003). In addition, Stuebing et al. (2009) evaluated the specific role of IQ as a predictor of intervention response. These reviews were illuminating in terms of BLCs that are related to reading performance and that potentially contribute to response to early intervention.

However, a meta-analytic synthesis of the literature that considers the analytic model and aggregates like effect sizes would facilitate an understanding of the magnitude of the role of specific cognitive skills in relation to intervention response. To this end, the first research question addressed whether BLCs relate to growth in reading over the course of intervention. The second research question examined the extent to which BLCs relate to growth that is not predicted by initial status. Finally, the third question was whether some measures were more predictive than others.

### Unconditional Relations Between Baseline Predictors and Intervention Response (Models 1 and 2)

The average fixed effect correlations between BLC and intervention response ranged from .12 to .60 for Model 1 and were significantly heterogeneous. However, the only BLC whose 95% CI did not include the mean meta-analytic effect of $r$ = .31 was oral language, based on results from five studies. The largest and smallest effects in this set (print knowledge, $r$ = .12; reading comprehension, $r$ = .60) were based on a single study and thus had the most sampling error.

The most defensible estimate of the relation between BLC and response to intervention defined as growth curves for all BLCs other than oral language is $r$ = .33, the mean effect when oral language effects are deleted because this set of effects is not significantly heterogeneous. For the most part, interpreting the observed differences is not warranted and do not really support differences in relative predictiveness. On average, the BLCs in this set account for 9% to 11% of the variance in response to intervention defined as growth curves (depending on whether we use the mean $r$ = .31 or $r$ = .33). For Model 2, the average fixed effect correlations across BLCs ranged from $r$ = .07 to $r$ = .33, with all of the observed variance in effects being consistent with sampling error. On average, BLCs accounted for 4% of the variance in response to intervention defined as gain. Although much is made of which skills are most predictive of intervention response, Models 1 and 2 do not show major differences across predictors failing to provide support for the hypotheses under the third research question.

### Unique Correlation Between BLCs and Gain (Model 3)

The average fixed effect (semipartial correlation) over all BLCs in predicting gain conditioned on initial status was $r$ = .15, with individual BLCs ranging from $r$ = −.04 (fluency) to $r$ = .33 (reading comprehension). However, the test of homogeneity indicated that this observed variability was consistent with sampling error and the estimated true variance of the effects was 0. The average effect for this model suggests that the unique variance predictable from BLCs when pretest is controlled is approximately 2% of the variance in posttest. Interestingly, this finding is comparable to the results for IQ in the Stuebing et al. (2009) meta-analysis.

The mean association between BLC and intervention response within the conditional model is weaker than that found in the unconditional models. This difference makes sense because it would be rare for the bivariate correlation between two variables to equal the semipartial correlation between the same two variables, when a third variable is partialed from one of them unless the third variable is uncorrelated with the other two. However, the BLCs in this meta-analysis were chosen because of demonstrated associations with reading in past studies. Thus, we would expect all BLCs to correlate with the Time 1 measure of reading.

## Consistency of Results With Previous Research

As anticipated, our effect sizes are generally smaller than those reported for the same BLCs in the meta-analysis by Scarborough (1998). Those studies were based on unselected samples of children and would not be subject to the restriction of range expected in intervention studies in which students are selected for poor reading. The BLCs were also used to predict reading at a later point in time, not growth in reading, although for novice readers, growth during the 1 to 2 years after the beginning of reading instruction will be highly correlated with performance at that posttest. Our effects are also smaller than those reported by Swanson et al. (2003), where BLCs were correlated with measures of reading assessed at the same time point. For BLCs that had five or more effects and thus relatively stable estimates, measures of PA and RAN were generally the best predictors, although not significantly different from one another and not really different from measures of letter-word identification or spelling. Our effects tended to be smaller than those found for the same BLCs by Nelson et al. (2003). To try to understand the differences, all seven studies that reported effects of Rapid Naming were located. Each one was coded according to predictor, outcome, population, time sampling, and analytic model. Additional study factors that could increase or decrease the observed correlation were also noted.

Although we cannot be absolutely certain that we have coded exactly the effects coded by Nelson et al. (2003), we arrived at virtually the same numeric value in six of the seven studies. In one study (Torgesen, Wagner, Rashotte, Rose, et al., 1999), we were not able to code an effect because only an unstandardized regression weight was available. Without knowing the standard deviation of the slope parameters in the model, we could not convert this into a correlation. We were able to estimate the correlation that would be just significant at the level claimed for this weight ($p < .001$), which was $r = .32$. Across these seven studies the models were mixed, which means the same effect was not being estimated. BLC was used to predict final status in three studies. In one of these (Berninger et al., 2002), the child characteristic was measured after the intervention rather than before. Effects from this model were not included in our current meta-analysis because this is not a model of change.

These effects are analytically similar to those found in Scarborough (1998) for RAN but on restricted samples. Gain was predicted from BLC in one study and growth was predicted from the BLC in the remaining three. These models parallel those coded in our study. Populations were mixed across studies. Some of the reported effects were based on mixtures of control and intervention students, some on just intervention, and in one case, just average children. In our study, we only coded effects for at risk students who had received intervention. Although the selection of subpopulations tends to reduce the observed correlation relative to what would be found if students across the entire range of abilities were selected, in several studies extreme groups were selected resulting in exaggeration of the effect size (Preacher, Rucker, MacCallum, & Nicewander, 2005). Finally, we note that the *N* reported by Nelson et al. (2003) in their Table 2 was the *N* for the whole sample and not for the effect. It is possible that if this *N* was used in calculating the weights for the effects, studies would be weighted inappropriately and the resulting mean would be inaccurate.

## Predicting Intervention Response: Unconditional or Conditional Growth Models?

None of the three growth models evaluated is better or worse than the alternatives. The value of the model depends on the question to be answered (Hand, 1994). The question answered by the unconditional growth models (Models 1 and 2) pertains to whether the BLC is associated with gain or growth, without investigating the cause of the association. In contrast, conditional growth models (Model 3) attempt to parse the associations between growth and initial status and a BLC. This model asks, “Does the BLC uniquely predict performance at posttest after controlling for pretest performance?” Similar to the unconditional growth models, the underlying cause of these associations is not evaluable. The association between the BLC and growth might be due to the fact that the BLC (a) causes change, (b) is caused by change (which is tenable in the Berninger et al. [2002] study where the BLC was measured after the intervention), or (c) because both growth and the BLC are caused by a variable not included in the model. However, by focusing on growth unexplained by pretest performance, the conditional growth model has valuable applications in educational settings. It allows us to make predictions about who will and will not respond to intervention among children reading at the same initial reading level. This is different from asking if children who have different reading status at Time 1 but who have the same level of the BLC grow at the same rate.

Different researchers in different contexts might prefer to ask one question or the other; no analysis is uniformly preferred. In early stages of research, it might be interesting to discover if there is any relation at all, before parsing the relation into parts. In later stages, it is more interesting to investigate if there is a direct effect of the BLC controlling for an indirect effect through the Time 1 assessment, although the causal model implied by this formulation must be plausible and is not provable via the analysis. Pragmatically, the researcher may want to know if the cost of administering an additional measure at Time 1 (the BLC) is offset by improvements in predictions of intervention effectiveness and thus allows for earlier adjustments in intervention content, duration, or intensity. Finally, although these effects are correlational, they help generate hypotheses regarding potentially malleable skills that can be targeted through intervention targeting. The key issue in the conduct and consumption of such research is clarity regarding the growth model utilized and the questions it can address.

## Predictors of Intervention Response

**Elaborating Theoretical Relations—**Care should be taken in interpreting observed differences in mean effects across individual predictors of intervention response. In different models, some predictors are slightly more correlated with change than others. However, in Models 2 and 3, the test for heterogeneity among effects was not significant. In Model 1, the only predictor that did not include the mean meta-analytic effect within its 95% CI (oral language) was based on a small number of studies. Furthermore, interpretation of the relation between unique predictors and growth is complicated because BLC measures are intercorrelated and unique relations with change are weak. In a dominance analysis predicting end of first-grade reading levels from beginning kindergarten cognitive assessments, Schatschneider et al. (2004) found that although phonological awareness was the best predictor of outcomes in that it shared unique variance with outcomes when the other predictors were controlled, it was also highly redundant with the other measures investigated. Working memory, for example, rarely accounts for unique variance in predicting reading skills, but this is because at the latent level, working memory is highly correlated with the other predictor variables (Wagner, 1996). It would be difficult to support conclusions that phonological awareness is more related to outcomes than working memory, especially because many phonological awareness assessments include a working memory component.

**Improving Screening Procedures—**The results of this meta-analysis suggest that predictors of intervention response vary dependent on the model of change used to measure response to intervention. However, differences among predictors within models were small and not statistically significant. The benefit of adding multiple cognitive variables would also likely be small, because the potential predictors are intercorrelated among themselves and also correlated moderately to pretest and posttest reading measures. If the sample is large enough, the addition of cognitive variables may be statistically significant, but the costs of administration relative to increased accuracy should be weighed.

To weigh the potential costs and benefits, it is important to understand the magnitude of the effects and their utility in real-world situations, especially for those effects observed in Model 3. The average effect for Model 3 was $r = .15$, which means that assessment of one of these BLCs would allow additional prediction of the posttest score over and above that allowed by the pretest of about 2.25% of the variance in posttest. It should be noted that this is the additional predictability relative to the entire variance of the posttest. We can illustrate the utilitarian impact of this additional 2.25% through an example. If, in our highly selected sample, the correlation between pretest and posttest is about .69 (Scarborough, 1998), we can account for 47.6% of the variance in the posttest with only the pretest. If adding a BLC improves the variance accounted for by 2.25%, the correlation between the posttest and the regression weighted linear composite of the pretest and BLC will be about .71, which when squared yields an $r^2$ of about .50. If we dichotomize two measured variables that have a correlation of .69 at their means, we are likely to get a cross tabulation of 1,000 observations similar to that seen in Table 4. If the correlation is improved to .71, the reduction in misclassifications is 8 out of 1,000 students tested equally split between false positives and false negatives (Table 5). With the same initial degree of correlation and a dichotomizing cut point at the 25th percentile (a likely cut point following intervention), adding 2.25% of predictive variance would reduce the misclassifications by 5 to 6 students out of 1,000. This estimate is based on using the mean estimated effect for all BLCs considered, which is justifiable considering the lack of heterogeneity among the tests.

However, even if we based our estimate of the improvement in prediction on the relatively high mean semipartial correlation of $r = .28$ for the seven RAN studies that account for an additional 7.4% of the outcome variance, the reduction in misclassifications per thousand students would be 26 per 1,000 for a cutoff at the median score and a reduction of about 19 if the split is made at the 25th percentile. Whether these relatively small reductions in misclassifications are an efficient use of resources depends on the cost of administering the assessments to all students as well as the availability of other potentially more efficient predictors. Overall, the evidence that assessments of specific cognitive skills have a major value added effect beyond the assessment of baseline reading skills is not compelling because all the effects are small. Because the goal of reading intervention is to improve reading, assessment of the BLCs involving reading (i.e., the pretest or assessment of initial status) may be sufficient.

## Limitations of the Study and Future Research

It is important to understand the results of this primarily descriptive study in terms of the limitations of the present synthesis of the literature on predictors of response. The limitations are presented with respect to the magnitude of the results, the inferences that may be made about them and limitations that pertain to interpretation and practice. Descriptively, the mean correlations within each of the three models are different and are ordered as we expected. We were not able to conduct an inferential test of the differences in the mean effects across the models due to the lack of independence across effects. The confidence intervals around each mean effect in Table 3 would suggest that the mean effects of Models 1 and 2 are different from each other, but we know that confidence intervals tend to be too narrow when there are dependencies among observations. The results are suggestive of the pattern we predicted, but we are not able to rigorously test for differences.

The magnitude of these mean correlations is most certainly attenuated by unreliability in the measurement of the constituent variables. In Models 1 and 2, we would expect the true correlations to be larger than the observed correlations if perfectly measured variables had been available. In Model 3, the effect of unreliability is more complex and depends on the intercorrelations among the pretest, the posttest, and the BLC. For example, a negative correlation is induced between gain and initial status when both the pre- and posttest are standardized but a positive correlation might be found if the measures exhibit fan spread. The true correlation could be larger or smaller depending on the particular pattern.

The confidence intervals of the mean correlations in Table 3 are correct because these effects all represent independent samples. However, it is not possible to perform an accurate inferential test of the difference between the means because there are dependencies across BLCs. To answer Question 3, we conducted tests of homogeneity to determine if there was variability across BLCs within models over and above what would be expected due to sampling. The observed differences were not significantly different in Models 2 and 3 and were homogeneous in Model 1 when oral language effects were removed.

The semipartial correlation coded for Model 3 studies represented a choice made by the authors. We chose to code the semipartial, which is in essence the correlation between the posttest and residual of the BLC when predicted from the pretest. Another choice would have been to code the partial correlation or the amount of growth in reading that was not predictable from initial status, but that was related to the part of the BLC that was also not related to initial status. In this case, the resulting correlation would be the association between the residuals of both posttest and BLC when initial status is controlled. This analysis might represent exactly the research question of interest to some, but because it is not directly comparable to the results from Models 1 and 2, we chose not to code it. The partial correlation will be as large or larger than the semipartial depending on the degree (but not direction) of correlation between the gain and initial status. Conceptually, it is important to understand that these correlations are potentially biased and might represent spurious relations if important explanatory variables have been omitted from the models. There is no statistical test to determine a misspecified model. It is based on the quality of the theoretical model of the causal effects among the variables and it is up to the consumer of these results to interpret them according to a defensible causal model.

This set of models is not exhaustive and there are other ways of asking the questions represented here that require different data collection and analytic approaches. In addition to the regression-based approach in Schatschneider et al. (2004), Connor et al. (2009) looked at the interaction of BLCs with intervention type, finding interactions of the type of reading deficit (decoding, comprehension) and the differentiation of instruction in these two domains based on strengths and weakness in these domains. Models of this sort are not assessed within the models in this meta-analysis.

A conceptual limitation of this study involves the pairing of BLCs with a variety of indicators of reading growth (e.g., accuracy, fluency, comprehension.). Within the results reported for each BLC, there is a mix of pairings. There were not enough effects within all possible pairs to analyze and report results separately. As a result, the results should be interpreted as the average effect for a given BLC over a range of possible growth outcomes, rather than as the single effect for the BLC. Although these limitations affect the interpretation of the results, they were not addressable in this study but rather represent that state of the existing literature. The results stand as a descriptive account of the current literature on correlates of RTI. That said, the limited heterogeneity in the effects within each class of models gives us confidence that the variability due to outcome choice is relatively small.

## Conclusions

The current meta-analysis examined the magnitude of the relations between baseline learner characteristics and response to intervention within each of three analytic models. Results indicated that the magnitude of the effects were consistent with our predictions, with the largest correlations found in Model 1 (growth curves), the second largest found for Model 2 (gain scores), and the smallest effects found when pretest was controlled (Model 3). Effects were homogeneous within Models 2 and 3 overall and also in Model 1 once oral language (which had an unusually small effect for this model class) was removed as a predictor. There was not strong evidence of major differences among predictors and the model of change may be more important than these differences in mean correlations.

Our effects tended to be much smaller than those found in previous meta-analyses of this literature. One of the reasons was that the early studies looked at unrestricted samples and were not subject to restriction of range, and also that the models used in these studies tended to be those that predicted final status from the BLC without taking initial status on reading into account. Even in the more recent Nelson et al. (2003) meta-analysis, it is possible that the effects from growth models were mixed with models of BLC predicting final status, thus inflating the overall mean estimated effect. If one of the uses for this literature is to select measures with high incremental predictive power, it is important that we derive estimates of their magnitude from studies that actually estimate that effect (Model 3).

The amount of variance added to prediction by all of the BLCs included in this study do not improve prediction in a clinically meaningful way and it is unlikely that administering tests of these characteristics to all students at risk for reading problems will be cost effective. As Vellutino et al. (2006) observed, there is little evidence that cognitive predictors beyond assessments of reading skills used at pretest have value added contributions to the prediction of intervention response. The exception is for preschool and kindergarten students who have not yet begun instructional programs. Here, a combination of different cognitive variables may predict those at risk as well as intervention response, although perhaps no better than measures with a letter component, such as letter identification and rapid letter naming. These tests may be simply early reading measures and not cognitive processes independent of the print component.

## Acknowledgments

**Note**

This research was supported in part by Grant P50 HD052117 from the Eunice Kennedy Shriver National Institute of Child Health and Human Development (NICHD). The content is solely the responsibility of the authors and does not necessarily represent the official views of the NICHD or the National Institutes of Health.

## APPENDIX

## Attention

- Multigrade Inventory for Teachers (Agronin, Holahan, Shaywitz, & Shaywitz, 1992): Torgesen et al. (2001).
- IOWA Conners Teacher Rating Scale (Loney & Milich, 1982): Torgesen et al. (2001).
- Researcher-Developed Measure—Attention Rating: Berninger et al. (1998).
- Swanson, Nolan, and Pelham (SNAP) rating scales (Atkins, Pelham, & Licht, 1985): Torgesen et al. (2001).

## Fluency

- Monitoring Basic Skills Progress (Fuchs, Hamlett, & Fuchs, 1990): Allor, Fuchs, and Mathes (2001).
- Test of Word Reading Efficiency (Torgesen, Wagner, & Rashotte, 1999): Case, Speece, and Molloy (2003); Torgesen et al. (2001).

## Nonverbal Processing

- Wechsler Intelligence Scale for Children–Revised (Wechsler, 1974): Hatcher and Hulme (1999); Vellutino et al. (2000).

## Oral Language

- British Picture Vocabulary Scales–Revised (Dunn, Dunn, Whetton, & Burley, 1997): Hatcher, Goetz, et al. (2006); Nash and Snowling (2006).
- Clinical Evaluation of Language Fundamentals–Third Edition (Semel, Wiig, & Secord, 1995): Torgesen et al. (2001).
- Expression, Reception, and Recall of Narrative Instrument (Bishop, 2003): Nash and Snowling (2006).
- Peabody Picture Vocabulary Test–Revised (Dunn & Dunn, 1981): Al Otaiba and Fuchs (2006); Vadasy, Sanders, and Abbott (2008); Vellutino et al. (1996).
- Picture Naming (Snowling, van Wagtendonk, & Stafford, 1988): Hatcher and Hulme (1999).
- Stanford-Binet: Fourth Edition (Thorndike, Hagen, & Sattler, 1986): Hecht and Close (2002); Torgesen and Davis (1996).
- Wechsler Intelligence Scale for Children–Revised (Wechsler, 1974): Hatcher and Hulme (1999); Vellutino et al. (2000).
- Wechsler Intelligence Scale for Children–III (Wechsler, 1991): Abbott, Reed, Abbott, and Berninger (1997); Berninger et al. (1998).

## Phonological Awareness

- Comprehensive Test of Phonological Processing (Wagner, Torgesen, & Rashotte, 1999): Hecht and Close (2002); Torgesen et al. (2001); Case et al. (2003).
- Lindamood Auditory Conceptualization Test (Lindamood & Lindamood, 1979): Wise, Ring, Sessions, and Olson (1997); Wise, Ring, and Olson (1999).
- Lindamood-Bell Auditory Conceptualization Test–Revised (Lindamood & Bell, 1989): Uhry and Shepherd (1997); Wise, Ring, and Olson (2000).
- Illinois Test of Psycholinguistic Abilities (Kirk, MacCarthy, & Kirk, 1968): Urhy and Shepherd (1997).
- Phonological Abilities Test (Muter, Hulme, & Snowling, 1997): Hatcher, Hulme, et al. (2006).
- Researcher-Developed Measure—Sound Discrimination: Hatcher and Hulme (1999).
- Researcher-Developed Measure—Segmenting: Berninger et al. (1999); Hatcher and Hulme (1999); O’Connor, Jenkins, and Slocum (2005).

- Researcher-Developed Measure—Phoneme and/or Syllable Deletion: Berninger et al. (1998); Berninger et al. (1999); Hatcher and Hulme (1999); O'Shaughnessy & Swanson (2000); Stage, Abbott, Jenkins, and Berninger (2003); Wise et al. (1997); Wise et al. (1999, 2000).
- Researcher-Developed Measure—Articulatory Awareness: Wise et al. (1997).
- Roswell-Chall Auditory Blending Test (Chall, Roswell, & Blumenthal, 1963): Uhry and Shepherd (1997).
- Sound Linkage Test of Phonological Awareness (Hatcher, 2000): Hatcher, Hulme, et al. (2006); Hatcher, Goetz, et al. (2006).
- Synthesis and Analysis Tests in the Torgesen-Wagner Battery (Wagner, Torgesen, & Rashotte, 1994): Foorman, Fletcher, Francis, Schatschneider, and Mehta (1998).
- Test of Phonological Awareness (Torgesen & Bryant, 1994): Allor et al. (2001); O'Shaughnessy and Swanson (2000); Torgesen and Davis (1996).
- Yopp Singer Test of Phoneme Segmentation (Yopp, 1995): Al Otaiba and Fuchs (2006); Vadasy et al. (2008).

## Print Knowledge

- Concepts about Print Test (Clay, 1979): Hecht and Close (2002).
- Researcher-Developed Measures—Print Conventions: Vellutino et al. (1996).

## Spelling/Orthographic Processing

- Peabody Individual Achievement Test–Revised (Markwardt, 1989): O'Shaughnessy and Swanson (2000).
- Researcher-Developed Measure—Orthographic Choice Task: Berninger, et al. (1999); Scheltinga, van der Leij, and Struiksma (2009); Stage et al. (2003).
- Researcher-Developed Measure—Orthographic Coding Task: Berninger et al. (1998); Berninger et al. (1999).
- Researcher-Developed Measure—Phonetic Spelling: Hatcher, Hulme, et al. (2006); Hatcher, Goetz, et al. (2006).
- Researcher-Developed Measure—Spelling Nonwords: Torgesen and Davis (1996).
- Wide Range Achievement Test (Wilkinson, 1995): Hecht and Close (2002); Vadasy et al. (2008).
- Wechsler Individual Achievement Test (Wechsler, 1992): Uhry and Shepherd (1997).

## Rapid Naming

- Comprehensive Test of Phonological Processing (Wagner et al., 1999): Case et al. (2003); Torgesen et al. (2001).

- Researcher-Developed Measure—Rapid Letter Naming: Al Otaiba and Fuchs (2006); Allor et al. (2001); Berninger et al. (1999); O'Connor et al. (2005); Stage et al. (2003); Uhry and Shepherd (1997).
- Researcher-Developed Measure—Rapid Digit Naming: Berninger et al. (1999); Scheltinga et al. (2009); Torgesen and Davis (1996); Uhry and Shepherd (1997).
- Researcher-Developed Measure—Rapid Object Naming: Uhry and Shepherd (1997).
- Researcher-Developed Measure—Rapid Color Naming: Allor et al. (2001); Berninger et al. (1999); Uhry and Shepherd (1997).
- Researcher-Developed Measure—Rapid Naming (Mixed Letters, Colors, and/or Digits): Allor et al. (2001); Wise et al. (1999); Wise et al. (2000).

## Reading Comprehension

- State or District Summative Assessments: Marr, Algozzine, Nicholson, and Dugan (2010).
- Woodcock Reading Mastery Test–Revised (Woodcock, 1987): Allor et al. (2001).
- Woodcock-Johnson Tests of Achievement–Revised (Woodcock & Johnson, 1989): Case et al. (2003).

## Word Reading

- British Ability Scales II (Elliott, Smith, & McCulloch, 1997): Hatcher, Hulme, et al. (2006); Hatcher, Goetz, et al. (2006).
- *Drie Minuten Toets* [Three Minute Test] (Verhoeven, 1995): Scheltinga et al. (2009).
- Researcher-Developed Measure—The Early Word Reading Test: Hatcher, Hulme, et al. (2006); Hatcher, Goetz, et al. (2006).
- Researcher-Developed Measure—Reading Nonwords: Torgesen and Davis (2000).
- San Diego Quick Assessment (LaPray & Ross, 1986): Fitzgerald (2001).
- Woodcock-Johnson Tests of Achievement–Revised (Woodcock & Johnson, 1989): Case et al. (2003); Hecht and Close (2002).
- Woodcock Reading Mastery Test–Revised (Woodcock, 1987): Allor et al. (2001); Al Otaiba and Fuchs (2006); Berninger and Traweek (1991); O'Shaughnessy and Swanson (2000); Torgesen et al. (2001); Uhry and Shepherd (1997); Vadasy et al. (2008); Vellutino et al. (1996); Vellutino et al. (2000).

## Working Memory

- Comprehensive Test of Phonological Processing (Wagner et al., 1999): Torgesen et al. (2001).

- Researcher-Developed Measure—Sentence Span Task: O'Shaughnessy and Swanson (2000).
- Wechsler Intelligence Scales for Children-III (Wechsler, 1991): Uhry and Shepherd (1997).

## References

References marked with an asterisk indicate studies included in the meta-analysis. References marked with a ^ indicate studies included in the small replication in the Discussion.

*. Abbott SP, Reed E, Abbott RD, Berninger VW. Year-long balanced reading/writing tutorial: A design experiment used for dynamic assessment. Learning Disability Quarterly. 1997; 20:249–263. doi:10.2307/1511311.

Agronin, ME.; Holahan, JM.; Shaywitz, BA.; Shaywitz, SE. The Multigrade Inventory for Teachers. In: Shaywitz, SE.; Shaywitz, BA., editors. Attention deficit disorder comes of age. PRO-ED; Austin, TX: 1992. p. 29-67.

Al Otaiba S, Fuchs D. Characteristics of children who are unresponsive to early literacy intervention: A review of the literature. Remedial and Special Education. 2002; 23:300–316. doi: 10.1177/07419325020230050501.

*. Al Otaiba S, Fuchs D. Who are the young children for whom best practices in reading are ineffective? An experimental and longitudinal study. Journal of Learning Disabilities. 2006; 39:414–431. doi:10.1177/00222194060390050401. [PubMed: 17004674]

*^. Allor JH, Fuchs D, Mathes PG. Do students with and without lexical retrieval weaknesses respond differently to instruction? Journal of Learning Disabilities. 2001; 34:264–275. doi: 10.1177/002221940103400306. [PubMed: 15499880]

Atkins MS, Pelham WE, Licht M. A comparison of objective classroom measures and teacher ratings of attention deficit disorder. Journal of Abnormal Child Psychology. 1985; 13:155–167. doi: 10.1007/BF00918379. [PubMed: 3973249]

^. Berninger VW, Abbott RD, Vermeulen K, Ogier S, Brooksher R, Zook D, Lemos Z. Comparison of faster and slower responders to early intervention in reading: Differentiating features of their language profiles. Learning Disability Quarterly. 2002; 25:59–76. doi:10.2307/1511191.

*. Berninger VW, Abbott RD, Zook D, Ogier S, Lemos-Britton Z, Brooksher R. Early intervention for reading disabilities: Teaching the alphabet principle in a connectionist framework. Journal of Learning Disabilities. 1999; 32:491–503. doi: doi:10.1177/002221949903200604. [PubMed: 15510439]

*. Berninger VW, Traweek D. Effects of a two-phase reading intervention on three orthographic-phonological code connections. Learning and Individual Differences. 1991; 3:323–338. doi: 10.1016/1041-6080(91)90019-W.

*. Berninger VW, Vaughan K, Abbott RD, Brooks A, Abbott SP, Rogan L, Graham S. Early intervention for spelling problems: Teaching functional spelling units of varying size with a multiple-connections framework. Journal of Educational Psychology. 1998; 90:587–605. doi: 10.1037/0022-0663.90.4.587.

Bishop, DVM. Expression, Reception and Recall of Narrative Instrument (ERRNI). Psychological Corporation; London, England: 2003.

Borenstein, M.; Hedges, LV.; Higgins, JPT.; Rothstein, HR. Comprehensive meta-analysis, Version 2. Biostat; Englewood, NJ: 2005.

Borenstein, M.; Hedges, LV.; Higgins, JPT.; Rothstein, HR. Introduction to meta-analysis. John Wiley; Oxford, England: 2009.

Bryant PE, MacLean M, Bradley L, Crossland J. Rhyme and alliteration, phoneme detection, and learning to read. Developmental Psychology. 1990; 26:429–438. doi:10.1037/0012-1649.26.3.429.

Byrne, B.; Fielding-Barnsley, R.; Ashley, L.; Larsen, K. Assessing the child's contribution to reading acquisition: What we know and what we don't know. In: Blachman, BA., editor. Cognitive and linguistic foundations of reading acquisition. Lawrence Erlbaum; Hillsdale, NJ: 1997. p. 265-285.

Carroll, J. Human cognitive abilities: A survey of factor-analytic studies. Cambridge University Press; New York, NY: 1993.

*. Case LP, Speece DL, Molloy DE. The validity of a response-to-instruction paradigm to identify reading disabilities: A longitudinal analysis of individual differences and contextual factors. School Psychology Review. 2003; 32:557–582.

Catts H, Adlof S, Weismer S. Language deficits in poor comprehenders: A case for the simple view of reading. Journal of Speech, Language and Hearing Research. 2005; 49:2782–2793. doi: 10.1044/1092-4388(2006/023).

Center Y, Wheldall K, Freeman L, Outhred L, McNaught M. An evaluation of Reading Recovery. Reading Research Quarterly. 1995; 30:240–263. doi:10.2307/748034.

Chall J, Roswell FG, Blumenthal SH. Auditory blending ability: A factor in success in beginning reading. Reading Teacher. 1963; 17:113–118.

Clay, M. Stones: The concepts about Print Test. Heinemann; Auckland, New Zealand: 1979.

Connor CM, Piasta SB, Fishman B, Glasney S, Schatsneider C, Crowe E, Morrison FJ. Individualizing student instruction precisely: Effects of child × instruction interactions on first graders' literacy development. Child Development. 2009; 80:77–100. doi:10.1111/j.1467-8624.2008.01247.x. [PubMed: 19236394]

Cooper, H. Synthesizing research: A guide for literature reviews. 3rd. Sage; Thousand Oaks, CA: 1998.

Darlington RB. Multiple regression in psychological research and practice. Psychological Bulletin. 1968; 69:161–182. doi:10.1037/h0025471. [PubMed: 4868134]

Decker S, Hale J, Flanagan D. Professional practice issues in the assessment of cognitive functioning for educational applications. Psychology in the Schools. 2013; 50:300–313. doi:10.1002/pits.21675.

Denton CA, Tolar TD, Fletcher JM, Barth AE, Vaughn S, Francis DJ. Effects of Tier 3 intervention for student with persistent reading difficulties and characteristics of inadequate responders. Journal of Educational Psychology. 2013; 105:633–648. doi:10.1037/a0032581. [PubMed: 25308995]

Dunn, LM.; Dunn, LM. The Peabody Picture Vocabulary Test–Revised. American Guidance Service; Circle Pines, MN: 1981.

Dunn, LM.; Dunn, LM.; Whetton, C.; Burley, J. The British Picture Vocabulary Scale: Second edition. NFER-Nelson; Windsor, England: 1997.

Elliott, CD.; Smith, P.; McCulloch, K. British Ability Scales II. NFER-Nelson; Windsor, England: 1997.

Fletcher, JM.; Lyon, GR.; Fuchs, LS.; Barnes, MA. Learning disabilities: From identification to intervention. Guilford; New York, NY: 2007.

Fletcher JM, Stuebing KK, Barth AE, Denton CA, Cirino PT, Francis DJ, Vaughn S. Cognitive correlates of inadequate response to intervention. School Psychology Review. 2011; 40:2–22. Retrieved from http://www.ncbi.nlm.nih.gov/pmc/articles/PMC3485697/.

Fletcher JM, Vaughn S. Response to intervention: Preventing and remediating academic difficulties. Child Development Perspectives. 2009; 3:30–37. doi:10.1111/j.1750-8606.2008.00072.x. [PubMed: 21765862]

*. Fitzgerald J. Can minimally trained college student volunteers help young at-risk children to read better? Reading Research Quarterly. 2001; 36:28–46.

*. Foorman BR, Fletcher JM, Francis DJ, Schatschneider C, Mehta P. The role of instruction in learning to read: Preventing reading failure in at-risk children. Journal of Educational Psychology. 1998; 90:37–55. doi:10.1037/0022-0663.90.1.37.

Fuchs D, Hale J, Kearns D. On the importance of cognitive processing perspective: An introduction. Journal of Learning Disabilities. 2011; 44:99–104. doi:10.1177/0022219411400019. [PubMed: 21383103]

Fuchs, LS.; Hamlett, C.; Fuchs, D. Monitoring basic skills progress. PRO-ED; Austin, TX: 1990.

Hale, J.; Fiorello, C.; Thompson, R. Integrating neuropsychological principles with RTI for comprehensive school-based practice. In: Arzubi, ER.; Mambrino, E., editors. A guide to neuropsychological testing for health care professionals. Springer; New York, NY: 2010. p. 229-261.

Hand DJ. Deconstructing statistical questions. Journal of the Royal Statistical Society (Series A). 1994; 157:317–356. doi:10.2307/2983526.

Hatcher, PJ. Sound linkage: An integrated programme for overcoming reading difficulties. 2nd. Whurr; London, England: 2000.

*. Hatcher PJ, Goetz K, Snowling MJ, Hulme C, Gibbs S, Smith G. Evidence for the effectiveness of the Early Literacy Support programme. British Journal of Educational Psychology. 2006; 76:351–367. doi:10.1348/000709905X39170. [PubMed: 16719968]

*. Hatcher PJ, Hulme C. Phonemes, rhymes, and intelligence as predictors of children's responsiveness to remedial reading instruction: Evidence from a longitudinal intervention study. Journal of Experimental Child Psychology. 1999; 72:130–153. doi:10.1006/jecp.1998.2480. [PubMed: 9927526]

*. Hatcher PJ, Hulme C, Miles JNV, Carroll JM, Hatcher J, Gibbs S, Snowling MJ. Efficacy of small group reading intervention for beginning readers with reading-delay: A randomized controlled trial. Journal of Child Psychology and Psychiatry. 2006; 47:820–827. doi:10.1111/j.1469-7610.2005.01559.x. [PubMed: 16898996]

*. Hecht SA, Close L. Emergent literacy skills and training rime uniquely predict variability in responses to phonemic awareness training in disadvantaged kindergartners. Journal of Experimental Child Psychology. 2002; 82:93–115. doi:10.1016/S0022-0965(02)00001-2. [PubMed: 12083791]

Hunter, JE.; Schmidt, FL. Methods of meta-analysis: Correcting error and bias in research findings. Sage; Newbury Park, CA: 1990.

Kearns D, Fuchs D. Does cognitively focused instruction improve the academic performance of low-achieving students? Exceptional Children. 2013; 79:263–290. doi:10.1177/0022219411400019.

Kendeou P, Savage R, van den Broek P. Revisiting the simple view of reading. British Journal of Educational Psychology. 2009; 79:353–370. doi:10.1348/978 185408X369020. [PubMed: 19091164]

Kirk, SA.; MacCarthy, JJ.; Kirk, WD. Illinois Test of Psycholinguistic Abilities. University of Illinois Press; Champaign: 1968.

LaPray, M.; Ross, R. The graded word list: Quick gauge of reading ability. In: Ekwall, EE., editor. Ekwall reading inventory. 2nd. Allyn & Bacon; Boston, MA: 1986. p. 28-29.

Lindamood, C.; Lindamood, P. Lindamood Auditory Conceptualization Test. Teaching Resources Corporation; Hingham, MA: 1979.

Lindamood, P.; Bell, N. Lindamood-Bell Auditory Conceptualization Test–Revised. Lindamood–Bell Learning Processes; San Luis Obispo, CA: 1989.

Loney, J.; Milich, R. Hyperactivity, inattention, and aggression in clinical practice. In: Wolraich, M.; Routh, DD., editors. Advances in developmental and behavioral pediatrics. Vol. 3. JAI Press; Greenwich, CT: 1982. p. 1213-2147.

Markwardt, FC. Peabody Individual Achievement Test–Revised. American Guidance Service; Circle Pines, MN: 1989.

*. Marr MB, Algozzine B, Nicholson K, Dugan KK. Building oral reading fluency with peer coaching. Remedial and Special Education. 2010; 32:256–264. doi:10.1177/0741932510362202.

Maxwell, SE.; Delaney, HD. Designing experiments and analyzing data: A model comparison perspective. Wadsworth Publishing; Belmont, CA: 1990.

Muter, V.; Hulme, C.; Snowling, MJ. Phonological Abilities Test. NFER-Nelson; Windsor, England: 1997.

Naglieri J. Do ability and reading achievement correlate? Journal of Learning Disabilities. 2001; 34:304–305. doi:10.1177/002221940103400403. [PubMed: 15503573]

*. Nash H, Snowling M. Teaching new words to children with poor existing vocabulary knowledge: A controlled evaluation of the definition and context methods. International Journal of Language

and Communication Disorders. 2006; 41:335–354. doi:10.1080/13682820600602295. [PubMed: 16702097]

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

Nelson JR, Benner GJ, Gonzales J. Learner characteristics that influence the treatment effectiveness of early literacy interventions: A meta-analytic review. Learning Disabilities Research & Practice. 2003; 18:255–267. doi:10.1111/1540-5826.00080.

*. O'Connor RE, Jenkins JR, Slocum TA. Transfer among phonological tasks in kindergarten: Essential instructional content. Journal of Educational Psychology. 1995; 87:202–217. doi: 10.1037/0022-0663.87.2.202.

*. O'Shaughnessy TE, Swanson HL. A comparison of two reading interventions for children with reading disabilities. Journal of Learning Disabilities. 2000; 33:257–277. doi: 10.1177/002221940003300304. [PubMed: 15505964]

Pashler H, McDaniel M, Rohrer D, Bjork R. Learning styles: Concepts and evidence. Psychological Science in the Public Interest. 2009; 9:105–119. doi:10.1111/j.1539-6053.2009.01038.x. [PubMed: 26162104]

Preacher KJ, Rucker DD, MacCallum RC, Nicewander WA. Use of the extreme groups approach: A critical reexamination and new recommendations. Psychological Methods. 2005; 10:178–192. doi: 10.1037/1082-989X.10.2.178. [PubMed: 15998176]

Rogosa DR, Brandt D, Zimowski M. A growth curve approach to the measurement of change. Psychological Bulletin. 1982; 92:726–748. doi:10.1037/0033-2909.92.3.726.

Scarborough H. Predicting the future achievement of second graders with reading disabilities: Contributions of phonemic awareness, verbal memory, rapid naming, and IQ. Annals of Dyslexia. 1998; 48:115–136. doi:10.1007/s11881-998-0006-5.

Schatschneider C, Fletcher J, Francis D, Carlson C, Foorman B. Kindergarten prediction of reading skills: A longitudinal comparative analysis. Journal of Educational Psychology. 2004; 96:265–282. doi:10.1037/0022-0663.96.2.265.

*. Scheltinga F, van der Leij A, Struiksma C. Predictors of response to intervention of word reading fluency in Dutch. Journal of Learning Disabilities. 2010; 43:212–228. doi: 10.1177/0022219409345015. [PubMed: 20032286]

Semel, E.; Wiig, EH.; Secord, W. Clinical Evaluation of Language Fundamentals-3. Psychological Corporation; San Antonio, TX: 1995.

Share DL, McGee R, Silva PD. IQ and reading progress: A test of the capacity notion of IQ. Journal of the American Academy of Child and Adolescent Psychiatry. 1989; 28:97–100. doi: 10.1097/00004583-198901000-00018. [PubMed: 2914843]

Snowling M, van Wagtendonk B, Stafford C. Object-naming deficits in developmental dyslexia. Journal of Research in Reading. 1988; 11:67–85.

Speece D, Ritchey K, Cooper D, Roth F, Schatschneider C. Growth in early reading skills from kindergarten to third grade. Contemporary Educational Psychology. 2004; 29:312–332. doi: 10.1016/j.cedpsych.2003.07.001.

*. Stage SA, Abbott RD, Jenkins JR, Berninger VW. Predicting early response to reading intervention from verbal IQ, reading-related language abilities, attention ratings, and verbal IQ-word reading discrepancy: Failure to validate discrepancy method. Journal of Learning Disabilities. 2003; 36:24–33. doi:10.1177/00222194030360010401. [PubMed: 15490889]

Stuebing KK, Barth AE, Molfese PJ, Weiss B, Fletcher JM. IQ is not strongly related to response to reading instruction: A meta-analytic interpretation. Exceptional Children. 2009; 76:31–51. [PubMed: 20224749]

Stuebing KK, Fletcher JM, LeDoux JM, Lyon GR. Validity of IQ-discrepancy classifications of reading disabilities: A meta-analysis. American Educational Research Journal. 2002; 39:469–518. doi:10.3102/00028312039002469.

Swanson HL, Trainin G, Necoechea DM, Hammill DD. Rapid naming, phonological awareness, and reading: A meta-analysis of the correlational evidence. Review of Educational Research. 2003; 73:407–440. doi:10.3102/00346543073004407.

Thorndike, RL.; Hagen, EP.; Sattler, JM. Stanford-Binet Intelligence Scale: Fourth edition. Riverside; Chicago, IL: 1986.

*. Torgesen JK, Alexander AW, Wagner RK, Rashotte CA, Voeller KKS, Conway T. Intensive remedial instruction for children with severe reading disabilities: Immediate and long-term outcomes from two instructional approaches. Journal of Learning Disabilities. 2001; 34:33–58. doi:10.1177/002221940103400104. [PubMed: 15497271]

Torgesen, JK.; Bryant, BR. Test of Phonological Awareness. PRO-ED; Austin, TX: 1994.

*^. Torgesen JK, Davis C. Individual difference variables that predict response to training in phonological awareness. Journal of Experimental Child Psychology. 1996; 63:1–21. doi:10.1006/jecp.1996.0040. [PubMed: 8812025]

Torgesen, JK.; Wagner, RK.; Rashotte, CA. Test of Word Reading Efficiency. PRO-ED; Austin, TX: 1999.

^. Torgesen JK, Wagner RK, Rashotte CA, Rose E, Lindamood P, Conway T, Garvan C. Preventing reading failure in young children with phonological processing disabilities: Group and individual responses to instruction. Journal of Educational Psychology. 1999; 91:579–593. doi: 10.1037/0022-0663.91.4.579.

*. Uhry JK, Shepherd MJ. Teaching phonological recoding to young children with phonological processing deficits: The effect on sight-vocabulary acquisition. Learning Disability Quarterly. 1997; 20:104–125. doi:10.2307/1511218.

*. Vadasy PF, Sanders EA, Abbott RD. Effects of supplemental early reading intervention at 2-year follow-up: Reading skill growth patterns and predictors. Scientific Studies of Reading. 2008; 12:51–89. doi:10.1080/10888430701746906.

*. Vellutino FR, Scanlon DM, Lyon GR. Differentiating between difficult-to-remediate and readily remediated poor readers: More evidence against the IQ-achievement discrepancy definition of reading disability. Journal of Learning Disabilities. 2000; 33:223–238. doi: 10.1177/002221940003300302. [PubMed: 15505962]

*^. Vellutino FR, Scanlon DM, Sipay ER, Small SG, Chen R, Pratt A, Denckla MB. Cognitive profiles of difficult-to-remediate and readily remediated poor readers: Early intervention as a vehicle for distinguishing between cognitive and experiential deficits as basic causes of specific reading disability. Journal of Educational Psychology. 1996; 88:601–638. doi: 10.1037/0022-0663.88.4.601.

Vellutino F, Scanlon D, Small S, Fanuele D. Response to intervention as a vehicle for distinguishing between children with and without reading disabilities: Evidence for the role of kindergarten and first grade interventions. Journal of Learning Disabilities. 2006; 39:157–169. doi: 10.1177/00222194060390020401. [PubMed: 16583795]

Vellutino F, Tunmer W, Jaccard J, Chen R. Components of reading ability: Multivariate evidence for a convergent skills model of reading development. Scientific Studies of Reading. 2007; 11:3–32. doi:10.1207/s1532799xssr1101_2.

Verhoeven, L. Drie Minuten Toets: Handleiding. CITO; Arnhem, Netherlands: 1995. [Three Minute Test: Manual]

Wagner, R. Expertise in reading. In: Lyon, R.; Krasnegor, N., editors. Attention, memory, and executive functions. Paul H. Brooks Publishing; Baltimore, MD: 1996. p. 139-156.

Wagner R, Torgesen JK, Rashotte CA. Development of reading-related phonological processing abilities: New evidence of bidirectional causality from a latent variable longitudinal study. Developmental Psychology. 1994; 30:73–87. doi:10.1037/0012-1649.30.1.73.

Wagner, RK.; Torgesen, JK.; Rashotte, CA. Comprehensive test of phonological processes. PRO-ED; Austin, TX: 1999.

Wechsler, D. Wechsler Intelligence Scale for Children–Revised. Psychological Corporation; New York, NY: 1974.

Wechsler, D. Wechsler Intelligence Scale for Children III. Psychological Corporation; San Antonio, TX: 1991.

Wechsler, D. Wechsler Individual Achievement Test. Psychological Corporation; San Antonio, TX: 1992.

Wilkinson, GS. The Wide Range Achievement Test. 3rd. Psychological Corporation; San Antonio, TX: 1995.

*. Wise BW, Ring J, Olson RK. Training phonological awareness with and without explicit attention to articulation. Journal of Experimental Child Psychology. 1999; 72:271–304. doi:10.1006/jecp.1999.2490. [PubMed: 10074381]

*. Wise BW, Ring J, Olson RK. Individual differences in gains from computer-assisted remedial reading. Journal of Experimental Child Psychology. 2000; 77:197–235. doi:10.1006/jecp.1999.2559. [PubMed: 11023657]

*. Wise BW, Ring J, Sessions L, Olson RK. Phonological awareness with and without articulation: A preliminary study. Learning Disability Quarterly. 1997; 20:211–225. doi:10.2307/1511309.

Woodcock, R. Woodcock Reading Mastery Tests–Revised. American Guidance Service; Circle Pines, MN: 1987.

Woodcock, RW.; Johnson, MB. Woodcock-Johnson Psychoeducational Battery–Revised. Riverside; Chicago, IL: 1989.

Wolf M, Bowers PG. The double-deficit hypothesis for the developmental dyslexias. Journal of Educational Psychology. 1999; 91:415–438. doi:10.1037/0022-0663.91.3.415.

Wolf, M.; O'Brien, B.; Adams, KD.; Joffe, T.; Jeffrey, J.; Lovett, M.; Morris, RD. Working for time: Reflections on naming speed, reading fluency, and intervention. In: Foorman, BR., editor. Preventing and remediating reading difficulties: Bringing science to scale. York Press; Timonium, MD: 2003. p. 355-379.

Yopp HK. A test for assessing phonemic awareness in young children. Reading Teacher. 1995; 49:20–29. doi:10.1598/RT.49.1.3.

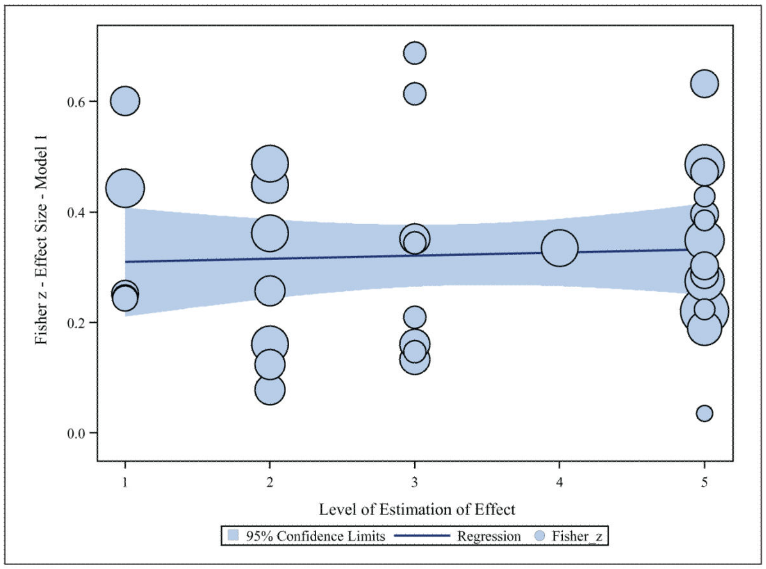


**FIGURE 1. Regression of effect size $Z_r$ on estimation—Model 1—Growth curves**

*Note*. Each symbol (i.e., circle) represents one effect or correlation between a baseline learner characteristic and the slope from a growth curve mode. Symbol size relates to the precision of the estimated effect with the diameter of the circle proportionate to the inverse variance of the effect. The regression line represents the linear relation between degree of estimation and the magnitude of the effect.

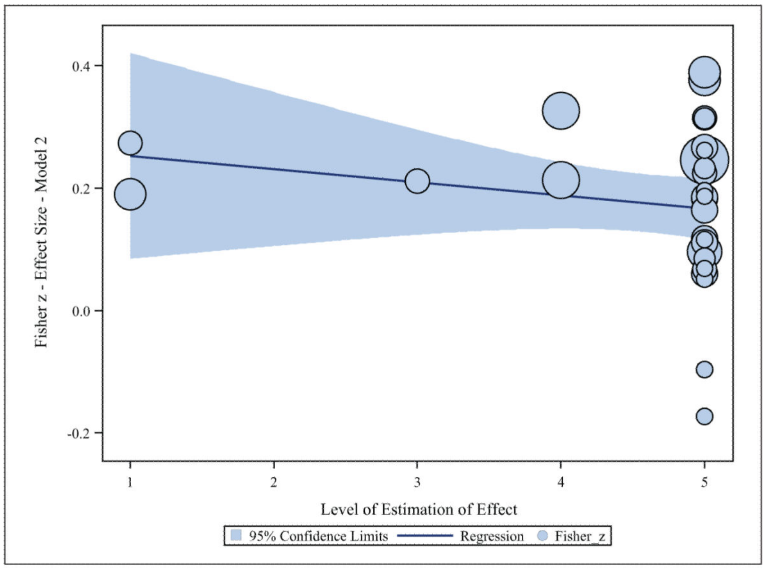


**FIGURE 2. Regression of effect size $Z_r$ on estimation—Model 2—Gain score**

*Note.* Each symbol (i.e., circle) represents one effect or correlation between a baseline learner characteristic and the gain between pretest and posttest. Symbol size relates to the precision of the estimated effect with the diameter of the circle proportionate to the inverse variance of the effect. The regression line represents the linear relation between degree of estimation and the magnitude of the effect.

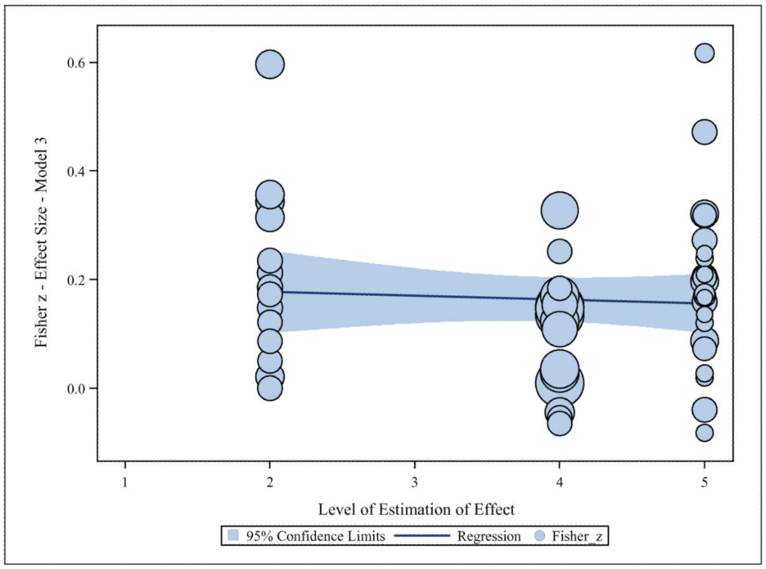


**FIGURE 3. Regression of effect size Zr on estimation—Model 3—conditional model**

*Note*. Each symbol (i.e., circle) represents one effect or the semipartial correlation between a baseline learner characteristic and a posttest controlling for the pretest. Symbol size relates to the precision of the estimated effect with the diameter of the circle proportionate to the inverse variance of the effect. The regression line represents the linear relation between degree of estimation and the magnitude of the effect.

## TABLE 1

Coding scheme for estimation of effects

| Code | Value labels | Example 1 | Example 2 |
| --- | --- | --- | --- |
| 1 | High estimation | High estimation Results include the statement that the effect was nonsignificant but no quantitative data were given. Low estimate was entered as 0 and high estimate was entered as the largest possible *r* that would be nonsignificant given the degrees of freedom for the test. | Results reported that the effect was significant, but no quantitative data or test statistics were reported. The low end estimate was the smallest possible significant effect given the degrees of freedom for the test. The high estimate was left missing, because there is no reasonable upper limit. |
| 2 | Moderate estimation | Beta weights for predicting response to intervention from initial ability and BLCs are reported but the correlations among the predictors are not included. $R^2$ and $r$ are computed by using correlations from large population studies as best guess for intercorrelations among predictors. | $F$ statistic or change in $R^2$ was reported, but the direction of the effect was unknown. We always coded a positive effect. When the population $r$ is positive, but small, a few negative effects will appear due to sampling error. Setting them all positive will lead to upward bias. |
| 3 | Some estimation | Degrees of freedom are not given for a particular analysis, and it is unclear how many additional covariates were included in the prediction equation, the sample size minus the estimated number of predictors was used to estimate degrees of freedom. When there are no missing data, this is not a problem, but if data are missing from variables included in the analysis and precise degrees of freedom are not given, the resulting effect size might be too small or too large. | Use of $\eta^2$ to estimate $r$. For example, the authors divide subjects into groups based on the amount of growth they showed in response to the intervention. The authors also reported the BLC means, standard deviations, and sample sizes for each group. With these data, it is possible to compute an $\eta^2$ for the linear relation between the ordered groups and the ability measure. We calculated sums of squares within, the sums of squares between, and the sums of squares due to the linear contrast only per Maxwell and Delaney (1990) and then formed the ratio of the sums of squares linear over the sums of squares total to arrive at an $\eta^2$ for the linear contrast. |
| 4 | Slight estimation | A $t$ statistic is given in the research report and well-known conversion formulae were used to convert the $t$ into an $r$. | The correlation matrix was given in the article and we input the matrix into SAS to compute the $R^2$ change between models. To the extent that the reported correlations are not as precise as raw data, this approach might result in a small amount of misestimation. |
| 5 | No estimation | $R^2$ change due to a predictor is given in the research report and the direction of the effect is also given. | Data were sent to our team by the authors of the study, and we analyzed it to obtain effect sizes. |

*Note.* BSL = baseline learner characteristics.

## TABLE 2

Effects by analytic group, aggregated and unaggregated

| Variable | *k* | *r* | 95% LCI | 95% UCI | *Q* | *df* | *P* | $I^2$ | τ |
|---|---|---|---|---|---|---|---|---|---|
| Growth curves | | | | | | | | | |
| Aggregated within study | 14 | .32[b] | 0.26 | 0.37 | 14.2 | 13 | .36 | 8.42 | .036 |
| Unaggregated | 36 | .31[a] | 0.27 | 0.36 | 51.84 | 35 | .03 | 32.48 | .083 |
| Gain score | | | | | | | | | |
| Aggregated within study | 10 | .22[b] | 0.14 | 0.3 | 2.03 | 9 | .99 | 0 | 0 |
| Unaggregated | 30 | .21[b] | 0.16 | 0.26 | 14.18 | 29 | .99 | 0 | 0 |
| Conditional growth model | | | | | | | | | |
| Aggregated within study | 18 | .15[b] | 0.08 | 0.23 | 5.28 | 17 | .99 | 0 | 0 |
| Unaggregated | 54 | .15[b] | 0.11 | 0.19 | 35.98 | 53 | .91 | 0 | 0 |

*Note.* LCI = lower limit confidence interval; UCI = upper limit confidence interval.

[a]Random effects meta-analytic mean.

[b]Fixed effects meta-analytic mean.

## TABLE 3

Meta-analytic mean correlations for separate BLC reported by analytic model

| | Growth curve = BLC | | | | Gain = BLC | | | | Post = Pre + BLC | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **BLC** | ***k*** | ***r*** | **LCI 95%** | **UCI 95%** | ***k*** | ***r*** | **LCI 95%** | **UCI 95%** | ***k*** | **r** | **LCI 95%** | **UCI 95%** |
| Attention | 1 | .42 | 0.25 | 0.56 | 1 | .18 | −0.11 | 0.44 | | | | |
| Fluency | 1 | .21 | −0.13 | 0.50 | 1 | .18 | −0.10 | 0.44 | 1 | −.04 | −0.39 | 0.33 |
| Nonverbal | 1 | .13 | −0.10 | 0.35 | | | | | 3 | .11 | −0.10 | 0.32 |
| Oral language | 5 | .19 | 0.08 | 0.30 | 1 | .12 | −0.17 | 0.39 | 8 | .13 | 0.02 | 0.23 |
| PA | 11 | .31 | 0.26 | 0.37 | 9 | .23 | 0.15 | 0.31 | 13 | .16 | 0.07 | 0.24 |
| Print knowledge | 1 | .12 | −0.11 | 0.34 | | | | | 1 | .30 | 0.00 | 0.56 |
| RAN | 5 | .34 | 0.24 | 0.43 | 7 | .18 | 0.08 | 0.29 | 7 | .27 | 0.15 | 0.38 |
| Reading Comp | 1 | .60 | 0.33 | 0.77 | | | | | 2 | .33 | 0.03 | 0.57 |
| Spelling | 5 | .38 | 0.29 | 0.46 | 2 | .33 | 0.13 | 0.51 | 6 | .14 | 0.03 | 0.24 |
| Word reading | 5 | .34 | 0.23 | 0.45 | 7 | .22 | 0.10 | 0.34 | 11 | .09 | 0.00 | 0.18 |
| Working memory | | | | | 2 | .07 | −0.19 | 0.32 | 2 | −.02 | −0.34 | 0.30 |

*Note.* No *Q* test of heterogeneity was significant for any baseline characteristic for any model. BLC = baseline learner characteristic; PA = phonological awareness; RAN = rapid automatized naming; Spelling = spelling and orthography; Nonverbal = nonverbal intelligence; Reading Comp = reading comprehension.

Author Manuscript

Author Manuscript

Author Manuscript

**TABLE 4**

Cross-tabulation of identification using pretest reading alone

| | Fail | Pass |
|---|---|---|
| Fail | 333 | 167 |
| Pass | 167 | 333 |

## TABLE 5

Cross-tabulation of identification using pretest reading and an additional predictor

| | **Fail** | **Pass** |
|---|---|---|
| Fail | 337 | 163 |
| Pass | 163 | 337 |