Journal of Educational Psychology

How You Read Affects What You Gain: Individual Differences in the Functional Organization of the Reading System Predict Intervention Gains in Children With Reading Disabilities

Noam Siegelman, Jay G. Rueckl, Mark van den Bunt, Jan C. Frijters, Jason D. Zevin, Maureen W. Lovett, Mark S. Seidenberg, Kenneth R. Pugh, and Robin D. Morris

Online First Publication, September 23, 2021. http://dx.doi.org/10.1037/edu0000672

CITATION

Siegelman, N., Rueckl, J. G., van den Bunt, M., Frijters, J. C., Zevin, J. D., Lovett, M. W., Seidenberg, M. S., Pugh, K. R., & Morris, R. D. (2021, September 23). How You Read Affects What You Gain: Individual Differences in the Functional Organization of the Reading System Predict Intervention Gains in Children With Reading Disabilities. *Journal of Educational Psychology*. Advance online publication. http://dx.doi.org/10.1037/edu0000672

Journal of Educational Psychology

ISSN: 0022-0663

https://doi.org/10.1037/edu0000672

How You Read Affects What You Gain: Individual Differences in the Functional Organization of the Reading System Predict Intervention Gains in Children With Reading Disabilities

2032

Noam Siegelman[1], Jay G. Rueckl[1, 2], Mark van den Bunt[1], Jan C. Frijters[3, 4], Jason D. Zevin[1, 5], Maureen W. Lovett[6, 7], Mark S. Seidenberg[8], Kenneth R. Pugh[1, 2], and Robin D. Morris[9]

[1] Haskins Laboratories, New Haven, Connecticut, United States
[2] Department of Psychological Sciences, University of Connecticut, Storrs Campus
[3] Department of Child and Youth Studies, Brock University
[4] Applied Disabilities Studies, Brock University
[5] Department of Psychology, University of Southern California
[6] Department of Paediatrics, University of Toronto, St. George Campus
[7] The Hospital for Sick Children, Toronto, Ontario, Canada
[8] Department of Psychology, University of Wisconsin–Madison
[9] Department of Psychology, Georgia State University

There is now considerable evidence regarding the types of interventions that are effective at remediating reading disabilities on average. It is generally unclear, however, what predicts the magnitude of individual-level change following a given intervention. We examine new predictors of intervention gains that are theoretically grounded in computational models of reading and focus on individual differences in the functional organization of the reading system. Specifically, we estimate the extent to which children with reading disabilities (n = 118 3rd–4th graders) rely on two sources of information during an oral word reading task—print–speech correspondences and semantic imageability—before and after a phonologically-weighted intervention. We show that children who relied more on print–speech regularities and less on imageability preintervention had better intervention gains. In parallel, children who over the course of the intervention exhibited greater increases in their reliance on print–speech correspondences and greater decreases in their reliance on imageability, had better intervention outcomes. Importantly, these two factors were differentially related to specific reading task outcomes, with greater reliance on print–speech correspondences associated with pseudoword naming, while (lesser) reliance on imageability related to word reading and comprehension. We discuss the implications of these findings for theoretical models of reading acquisition and educational practice.

Educational Impact and Implications Statement

Why do some children with reading disabilities show substantial gains when provided with an intervention program, whereas others show little to no improvement? Here we argue that such differences in improvement can be tied to the way in which different children read (reflecting the functional organization of their reading system). We measure how different children are impacted by two sources of information of words they read—print–speech correspondences (orthographic–phonological) and semantic imageability—before and after a phonologically-weighted intervention for children with reading disabilities. Results show that children who before the intervention relied more on orthographic–phonological knowledge, and less on semantic imageability, had greater reading skill gains over the intervention. Skill gains were also associated with changes children made in their use of these two sources of information

Noam Siegelman https://orcid.org/0000-0002-5005-4838
Jan C. Frijters https://orcid.org/0000-0001-5625-4892
Jason D. Zevin https://orcid.org/0000-0001-8929-5914
Robin D. Morris https://orcid.org/0000-0002-0784-8237

Research reported in this publication was supported by the Eunice Kennedy Shriver National Institute of Child Health & Human Development of the National Institutes of Health under Award P01HD070837 and P20HD091013. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health. Noam Siegelman received funding from the Israel Science Foundation (ISF), Grant 48/20.

Correspondence concerning this article should be addressed to Noam Siegelman, Haskins Laboratories, 300 George Street, New Haven, CT 06511, United States. Email: noam.siegelman@yale.edu

over the intervention. These findings may help in detecting the children who may benefit more or less from a given intervention and advance us toward identifying an optimal intervention program for each individual based on the state of their reading system.

Keywords: intervention response, reading disabilities, individual differences, print–speech regularities

Supplemental materials: https://doi.org/10.1037/edu0000672.supp

Over the past few decades, extensive research has been devoted to the evaluation—and improvement—of intervention methods aimed at the remediation of reading disabilities (RD). This extensive literature had a key role in establishing that certain classes of interventions are effective (e.g., National Reading Panel, 2000; National Research Council, 1998) and that some methods have better outcomes than others, highlighting in particular the positive effects of theory-based remediation programs that focus on explicit and systematic phonologically-based instruction (see Castles et al., 2018; Duff & Clarke, 2011, for reviews). Operationally, what is common to many of these studies is that they focus on group-level intervention effects—that is, the mean impact of a given intervention across a sample of participants and how it varies as a function of intervention type or relative to a control group (Galuschka et al., 2014; McArthur et al., 2012), as well as different intervention-related factors (e.g., number of intervention hours, intervention settings, grade in which the intervention is provided; e.g., Tran et al., 2011; Wanzek et al., 2013, 2016).

At the same time, it is also well-documented that not all individuals respond similarly to a given intervention. In fact, studies show that sometimes many children show very limited gains even when provided well-established intervention protocols (e.g., Torgesen, 2000 who estimated the proportion of low responders at 30%). A critical question therefore has to do with the predictors of individual-level skill gains—that is, what participant-level characteristics determine whether an individual will show more or less improvement over the course of a given remediation program. Such information can provide insights into why some children hardly respond to the same approach that works well for others, and may, in principle, help us to generate a priori hypotheses on what classes of approaches might work better for poor responders to canonical programs.

Broadly speaking, previous studies on individual differences in intervention gains have focused on three categories of predictors. The first comprises general individual-level cognitive characteristics. Thus, previous studies have linked response to intervention to IQ levels (e.g., Frijters et al., 2011; Lovett et al., 2017) and sequential learning abilities (e.g., van der Kleij et al., 2019). The second category tied response to intervention to well-established linguistic correlates of reading, such as performance in tasks measuring phonological awareness (e.g., Torgesen et al., 1999); vocabulary skills (e.g., Lovett et al., 2017); and rapid naming (e.g., Scheltinga et al., 2010). Finally, some studies examined the predictive power of basic early reading skills (reflecting severity of reading difficulties at baseline), including preintervention word reading skills (e.g., Vaughn et al., 2020). Generally speaking, meta-analyses and systematic reviews of predictors of individual-level responsiveness have likewise focused on these three categories of measures (e.g., Al Otaiba & Fuchs, 2002; Stuebing et al., 2015). Such meta-analytic studies also show that measures from these three broad categories generally account for limited variance in treatment response.[1]

The current article reports data from a large-scale treatment study using an evidence-based intervention that centers on phonologically-based decoding strategies (adaptation of the PHAST program; Lovett, Lacerenza, & Borden, 2000; see details under Method: The Intervention Protocol) to examine potential novel predictors of relative skill gains. Specifically, we examine the relation between a child's preintervention reading system's functional organization (as characterized by a computational model of reading) and their skill gains during intervention. Thus, instead of correlating the level of intervention skill gains with a child's general cognitive characteristics, their overall reading skills at baseline, or a particular correlate of reading ability, we examined the relations between the way in which a child reads, as reflected in their relative reliance on two different sources of information during an oral word reading task, and their skill gains. In other words, we asked how does an individual's current state of the reading system affect what they can learn from the additional experience provided as part of the intervention. We also examined how the functional organization of an individual's reading system changes during an intensive intervention, and whether individual reading gains are associated with individual-level changes in their reliance on these different sources of information.

The theoretical foundations of the current study are provided by a framework that holds that word reading draws on multiple sources of information and that differences in the relative contribution of these sources have systematic behavioral consequences. These assertions are axiomatic principles of the "triangle" model of reading (Seidenberg & McClelland, 1989), a computational theory positing that reading acquisition entails the formation of associations between the representations of the orthographic (O), phonological (P), and semantic (S) properties of written words. Beginning readers already have extensive knowledge of the associations between phonology and semantics after years of exposure to speech, yet the introduction of print requires them to learn new correspondences between orthography and phonology (O-P) and between orthography and semantics (O-S; see, e.g., Chang et al., 2019). Importantly, the O-P and O-S mappings differ in their statistical properties as they involve regularities with differing degrees of consistency or predictability. Thus, in alphabetic languages, O-P regularities are generally more consistent than O-S associations. As a result, O-P regularities are typically acquired earlier in development than O-S associations due to their more systematic nature (e.g., Harm & Seidenberg, 2004).

A large number of empirical studies have investigated the role of O-P and O-S computations in word recognition. Studies show that readers are faster and more accurate when reading words that are

[1] Note that much of the current literature focuses on somewhat limited age ranges (typically, on young readers). It is plausible that predictors of skill gain or intervention effects interact with the recipients' age (i.e., some predictors may be more pronounced early on in development, whereas some may be more important among older individuals).

O-P consistent (e.g., *mint* is read faster and more accurately than *pint*; e.g., Glushko, 1979; Jared et al., 1990), and that consistency effects are larger in less frequent words (e.g., Seidenberg, 1985). In parallel, characteristics of word meaning that are thought to impact the strength of O-S associations also influence word reading. In particular, readers recognize faster and more accurately words that are higher in imageability, defined as the ease of eliciting a mental image (e.g., *mint* is recognized faster and more accurately than *hope*; e.g., Coltheart et al., 1988; Klose et al., 1983). Studies also show that imageability effects are stronger in infrequent, O-P inconsistent words (for skilled readers, at least; e.g., Strain et al., 1995). This is interpreted as evidence for a division of labor between the O-P and O-S(-P)[2] components: Words that are harder to learn via the O-P pathway end up with more involvement from the indirect O-S(-P) pathway (Harm & Seidenberg, 2004; Plaut et al., 1996).

Of particular importance for the present study is that according to the model, individual differences in learning capacities (including in particular statistical learning abilities; see Sawi & Rueckl, 2019) and/or learning experiences can differentially affect the acquisition of knowledge of O-P and O-S regularities. Thus, not all readers are expected to rely similarly on the different regularities available to them in the written input (see, e.g., Harm & Seidenberg, 1999; Plaut et al., 1996; Rueckl, 2016; Rueckl et al., 2019; Woollams et al., 2007; Zevin & Seidenberg, 2006; for computational insights and discussion). In the context of emerging reading skills, this suggests that some early readers may be more sensitive than others to the different types of regularities in their writing system. Indeed, behavioral studies show that not all individuals display similar reliance on O-P and O-S pathways, and that these individual differences in code utilization weighting have important implications for understanding variability in reading skills across different developmental stages. Thus, earlier studies with adult populations provided evidence that the magnitude of O-P and imageability effects vary across individuals, and that poorer readers display greater imageability effects (e.g., Pugh et al., 2008; Strain & Herdman, 1999; Woollams et al., 2016; and see Davies et al., 2017, for evidence from both adults and children).

These results were recently extended in work by Siegelman, Rueckl, et al. (2020) that focused on individual differences in code utilization and their relation to reading skills among a group of 399 primary school-age children. Operationally, to estimate the extent to which a reader relies on O-P and O-S processes, this study used a word reading task in which participants were asked to read single words aloud (modeled after the earlier group-level work by Strain et al., 1995 and adapted for children), with a manipulation of items' O-P regularities and their imageability. Logistic models predicting item-level naming accuracy on the data from each reader resulted in two slope scores for each child: the individual-level effect of O-P regularities, and that of imageability. The first reflects the extent to which a child relies on O-P regularities—a source of information that in English (and in other alphabetic languages) provides a generally valid cue to access a word's phonology from its orthographic form. The second slope score—the extent to which a child relies on imageability—reflects the extent to which a child relies on an arbitrary semantic cue, and thus quantifies inefficient reliance on O-S processes. Siegelman, Rueckl, et al. (2020) found that individuals who rely more on O-P regularities during the word naming task, and less on imageability, had better reading skills as reflected in standardized reading test scores. These results were taken to suggest that children who rely more on efficient source of information (O-P regularities) and less on arbitrary cues (words' imageability) are better readers.

The Current Study

These computational and behavioral findings raise two questions with regard to individual-level variability in intervention gains among struggling readers. The first is how does the functional organization of the reading system—as reflected in the relative reliance on O-P regularities and imageability—change as a result of an intensive intervention. That is, we ask whether participants' mean reliance on O-P and imageability changes from pre- to postintervention, and whether individual differences in the level of intervention outcomes are associated with individual differences in changes in their reliance on these two sources of information. The second—and central—question is whether individual differences in preintervention reliance on O-P versus imageability are related not only to concurrent reading skills but also to skill gains following a phonologically-based decoding strategies intervention program: That is, how what the child knows already about their writing system affects what they learn from the intervention.

Several contrasting hypotheses can be made regarding this second question. One possibility is that an intervention centered on phonologically-based decoding strategies would be more effective for children who already show some (if underdeveloped) reliance on O-P regularities as they can continue to build upon this partial knowledge to achieve better reading skills. Alternatively, one could hypothesize that a phonologically-weighted program may be less beneficial for these children as they already have sufficient knowledge of print–speech correspondences. Under this hypothesis, it is the children who show less preintervention reliance on O-P who would show relatively greater intervention gains. Contrasting predictions can be similarly made regarding preintervention reliance on imageability: One could posit that a phonologically-weighted intervention would be less effective for children who are already strongly reliant on imageability as they seem to follow an atypical developmental trajectory; alternatively, perhaps those individuals would benefit more from the intervention as it could help them develop sensitivity to another cue that represents a more efficient source of information.

The current study addresses these questions by examining the relations between measures of preintervention and postintervention reliance on O-P and imageability during word reading and individual differences in skill gains in a large sample of children with RD who undergo an intense, phonological decoding strategies intervention. To preview our main findings, we show that (a) preintervention reliance on O-P regularities and imageability both predict intervention gains, but at the same time (b) each of these measures is predictive of gains in specific types of reading skills but not others.

[2] We use the label O-S(-P) (rather than O-S) to refer to the full indirect pathway from O to P via S, which involves not only mapping of orthography to semantics but also of semantics to phonology.

Method

Participants

Children in the third and fourth grade were recruited to an intervention study for children with RD, which included behavioral and neural measures before, during, and after an intensive phonologically-weighted intervention. Here we only report data from the RD subgroup that participated in the intervention (a smaller group of typically developing children, not receiving the intervention but serving as a comparison group, are not reported here; but see Siegelman, Rueckl, et al., 2020). We also only include children who have both pre- and postintervention data on the word reading task and full reading assessment data from four time points during the intervention (see below). This sample includes 118 children (67 boys; mean age in years: 9.29, *SD* = .69, range: 8.15 – 11.29). Data from additional $n = 21$ children with RD were excluded due to attrition ($n = 13$) and missing data ($n = 8$). Participants come from public and charter schools from a large urban community (74% African American, 20% Caucasian, 6% other/biracial). All children met an RD study-based criterion, defined as scoring at least one standard deviation below age-norm expectations on at least one standard reading assessment (see also Arrington et al., 2019). Participants were therefore characterized by below-average to low-average preintervention standard reading scores (Woodcock–Johnson III subtests, Woodcock et al., 2001: Letter–Word Identification: M = 88.42, SD = 8.30; Word Attack: M = 88.08, SD = 9.11; Passage Comprehension: M = 80.78, SD = 8.55; TOWRE-II, Torgesen et al., 2012: Sight Word Efficiency: M = 76.26, SD = 11.05; Phonemic Decoding Efficiency: M = 73.47, SD = 9.40). Note that preintervention data on a large portion of this sample (101/118 participants) are also included in the recent paper by Siegelman, Rueckl, et al. (2020) described above, which examines the concurrent (preintervention) associations between the measures from the word reading task and reading skills.

The Intervention Protocol

All children in the intervention study were placed into a small instructional group of between 5 and 8 students with similar word reading levels. These groups received 70 hr of intervention (typically 45–60 minutes daily) during the school year, as an add-on to their regular language arts and reading curriculum. Experienced research teachers implemented the intervention program for each group of children. The intervention was an adaptation of the PHAST Reading Program (described and evaluated in separate publications; Lovett et al., 2005; Lovett, Borden, et al., 2000; Lovett, Lacerenza, & Borden, 2000; Morris et al., 2012). The PHAST intervention was designed to remediate decoding deficits and teach effective word identification strategies. The program used direct instruction of letter–sound and letter–cluster correspondences, and phonological analysis and blending training, in addition to teaching a set of four word-identification strategies that emphasized larger orthographic patterns (e.g., onset-rime), awareness of the morphological structure of words, and variant vowel pronunciations.

To ensure more automatic word recognition, repeated and varied practice reading word parts (phonemes, rimes, affixes, etc.), and whole words (irregular words, high frequency key words), were included. The program also used complex multisyllabic words as "challenge" words on which the children practiced their strategies. Children who previously struggled to read one-syllable words independently worked, with support, through the decoding of multisyllabic words such as *unintelligible*, *comprehensive*, and *unrelentingly*, and in so doing considered themselves ready for what they called "college words." The PHAST approach combined direct and dialogue-based instruction, explicitly teaching children different levels of subsyllabic segmentation, and training them in the acquisition and effective use of multiple decoding strategies. The inclusion of strategy instruction and the promotion of a flexible approach to word identification were considered important to facilitate transfer of learning to uninstructed materials and texts.

Approximately 75%–80% of program activities were devoted to decoding and word identification work, with some additional time allocated to activities supporting text reading and reading comprehension. Twice a week, lessons included vocabulary work introducing one or two vocabulary words from texts being read that week; word webbing and a vocabulary chart were constructed in this component. While text reading was a reliable part of every lesson so that decoding skills and strategies were applied and practiced on words in context, formal comprehension instruction was added only for the 44th–70th lessons. This component included introduction of some comprehension strategies; students, for example, assessed their ongoing comprehension of text, and stopped to clarify sources of confusion before proceeding on. A graphic organizer was displayed to guide readers' activities before (ask questions, make predictions), during (e.g., identify characters, setting, time, problem in a narrative text), and after reading a text (summarizing). Only 5–10 minutes per lesson were allotted to the comprehension activities, however, because of the decoding focus of the lessons.

Fidelity of Implementation

Multiple teachers provided the interventions at several schools. Teachers had a detailed Scope and Sequence that provided specific (scripted) details for each day's lesson. A senior, experienced research teacher met each week with the teachers to monitor the progress of each instructional group through the lessons, answer questions and problem-solve around the program, and provide instructional refreshers. To further support fidelity of implementation, the senior research teacher randomly selected instructional groups to visit and monitor on a monthly basis during the program (more frequently for new teachers learning the program during their first year), observe teacher-student interactions, and observe immediate corrective feedback, providing the teachers feedback on their instruction. This model was designed to monitor that the teacher was providing the program as designed, introduced skills and strategies as planned (e.g., direct explanation vs. modeling), and used the appropriate support materials (e.g., posters, keyword bank) as proscribed by the program. In addition, every teacher completed daily lesson diaries for each child and each group. This diary reported which lessons and specific activities were completed, and whether there were any challenges in their implementation. The senior research teacher reviewed these and provided guidance and feedback as needed.

Word Reading Task: Materials, Design, and Procedure

Within a larger battery of behavioral, psychometric, and neuroimaging assessments, each child participated in an oral word reading task before and after the 70 hr of intervention. This task included 160 trials presented to children in a fixed order. On each trial, a fixation cross appeared for 500 ms and was then replaced by a monosyllabic word that was presented in the center of the screen until response. Participants were asked to read out loud each word as accurately and quickly as possible. All responses were double coded by two members of the research team. First, children's responses were coded (as correct/incorrect) by the examiner, who sat in the room (who could also relisten to the session's recordings if needed). Then, a second member recoded responses using the session's recordings. In cases of disagreement between coders, they would each relisten to the recordings and discuss until resolution.

Importantly, words varied along three independent variables: frequency, imageability, and O-P regularity (conceptually modeled after Strain et al., 1995). Frequency (log-transformed) was estimated for each word based on the Zeno corpus, grades 1–8 (Zeno et al., 1995), and words' imageability was based on standard ratings (Paivio et al., 1968). O-P regularity was operationalized as vowel surprisal (i.e., $-\log(p(i))$) of the vowel pronunciation, which is a function of the relative likelihood of the pronunciation of a vowel grapheme (estimates taken from Siegelman, Kearns, et al., 2020). Hence, for example, the word *pint* has a higher surprisal value than *mint*, since p(*i* →/aɪ/) is smaller than p(*i* →/ɪ/). Items were selected in a way that minimizes the correlations among the three independent variables across the 160 items (O-P surprisal and imageability: $r = -.08$; O-P surprisal and frequency: .04; imageability and frequency: .16). Accuracy in each trial was coded as 1 (correct) or 0 (incorrect). Trials with microphone malfunctions or unclear responses were disregarded from further analysis (13.3% of all trials preintervention; 11.1% postintervention). Note that accuracy in this task showed high internal consistency, with split-half reliability (Spearman-Brown corrected) estimated at .96. Please refer to Siegelman, Rueckl, et al. (2020) for a full list of items and their frequency, imageability, and O-P surprisal values.

Analysis of the Word Reading Task

Following Siegelman, Rueckl, et al. (2020), we first extracted slope scores from the word reading task that reflect reliance on O-P and imageability in reading aloud. To do so, we ran two logistic regression models on the accuracy data of each participant at each time point (pre- and postintervention), including the subject's trial-by-trial data from the word reading task. The two models used accuracy on each trial as the dependent variable. One model included O-P surprisal as the independent variable; while the second model included imageability as the independent variable.[3] Note that in both models the independent variable was centered and scaled. The slope of the first model was used as a proxy of the impact of O-P regularities on naming accuracy, and the estimated slope of the second model was used to estimate the individual-level impact of imageability.[4] Mean raw accuracy on the task ranged between 17.9% and 90.6% preintervention (M = 68.5%, SD = 14.8%), and between 25.0% and 92.4% postintervention (M = 73.7%, SD = 13.1%). Hence, within this RD sample, no children were at complete ceiling or floor, and it was thus possible to extract slope scores from the accuracy data of all children in this sample (see Siegelman, Rueckl, et al., 2020; for details regarding exclusion of subjects in this task).[5] Note that although in the O-P surprisal measure higher values represent more surprising (i.e., unpredictable) pronunciations, for simplicity we "flipped" the O-P slopes in all further analyses. Thus, slope scores of sensitivity to O-P regularity were coded such that higher, positive values represent more reliance on O-P regularities, in the same direction as the slope scores reflecting sensitivity to imageability. Overall, then, each child had four slope scores from the word reading task: two slope scores reflecting reliance on O-P and imageability preintervention; and two slope scores reflecting postintervention reliance on the two processes. In the analyses below, we use as predictors the preintervention reliance on O-P and imageability, as well as the change in O-P and imageability from pre- to postintervention (i.e., post- minus pre-O-P; post- minus preimageability).

Reading Assessment

In addition to the oral word reading task, each participant also completed reading assessments using the Woodcock-Johnson III Tests of Achievement battery (Woodcock et al., 2001) and TOWRE-II (Torgesen et al., 2012) at four time points over the course of the intervention (after 0, 23, 45, and 70 hr of instruction). Below we report data on three subtests of the Woodcock-Johnson battery: Letter-Word Identification (measuring letter and word decoding), Word Attack (pseudoword reading), and Passage Comprehension, as well as on Sight Word Efficiency and Phonemic Decoding Efficiency from TOWRE.[6] We use measures from both Woodcock-Johnson and TOWRE, as the former uses untimed tests (measuring accuracy) while the latter examines timed performance (also measuring efficiency/fluency). Note that all these measures are known to have high reliability (reliability estimates for all five subtests used here

[3] As noted above, the task also included a manipulation of word frequency. However, at this age range and population, reliance on frequency does not explain almost any variance beyond reliance on O-P and O-S (Siegelman, Rueckl, et al., 2020). Moreover, from a theoretical perspective, reliance on frequency is not specific to reliance on O-P vs. O-S (-P) pathways. Therefore, we only focus on these two slope scores as variables of interests.

[4] An alternative procedure is to use one model per participant including both predictors. Note, however, that items were selected to the task in a way that minimizes the correlations among independent variables. Thus, separate models and one simultaneous model are expected to result in very similar estimates of slope scores. Indeed, re-running our analysis using one model per subject (with both O-P surprisal and imageability) resulted in slope scores that were almost perfectly correlated with the estimates from the two separate models: $r = 0.993$ for reliance on imageability; $r = 0.995$ for reliance on O-P regularities.

[5] Following Siegelman, Rueckl, et al. (2020), we a priori decided not to analyze latency data from the reading task and focus on accuracy. This was done because participants in this age range tend to produce a large number of responses with unclear or inaccurate speech onsets, limiting the reliability of latency data.

[6] The assessment battery also included a measure of Reading Fluency from the Woodcock-Johnson. Here we opted not to report data from this subtest because it focuses on passage fluency, an advanced skill thar is most likely unstable at this sample of struggling young readers. Indeed, gains in this task over the course of the intervention were generally not correlated with other reading skills, which may suggest that improvements in this task are unreliable in the current sample.

previously estimated at or above .87; see Woodcock et al., 2001 and Torgesen et al., 2012).

Results

Preintervention Associations of O-P and Imageability With Reading Skills

As mentioned above, a large portion (n = 101/118) of the current sample is included in the recent work by Siegelman, Rueckl, et al. (2020) that reports preintervention correlations between the O-P and imageability measures from the word reading task and standardized measures of reading skill. For brevity, we do not repeat these analyses here: These preintervention associations can be found in the online supplemental materials. As expected, given the previous study, these analyses show that better reading skills (across subtests of Woodcock-Johnson and TOWRE) at preintervention were associated with greater reliance on O-P and lesser reliance on imageability.

Mean Changes in Reliance on O-P and Imageability as a Result of the Intervention

Before turning to the main focus of the paper—analyses of individual differences—we first report the mean changes in reliance on O-P and imageability from pre- to postintervention to examine whether a phonologically-weighted multicomponent intervention impacts the functional organization of the reading system at the group level. Perhaps unsurprisingly given the treatment focus, we found that reliance on O-P was significantly stronger postintervention compared to preintervention: $M_{pre} = .31$, $SD_{pre} = .21$; $M_{post} = .39$, $SD_{post} = .20$; $t(117) = 4.39$, $p < .001$, $d = .39$. There was no significant mean change in the impact of imageability, despite some numeric decrease: $M_{pre} = .33$, $SD_{pre} = .21$; $M_{post} = .29$, $SD_{post} = .20$; $t(117) = -1.48$, $p = .14$, $d = .17$. These results show that the intensive phonologically-weighted intervention resulted in an increased sensitivity to sublexical O-P regularities at the group level.

Estimating Individual Differences in Intervention Skill Gains

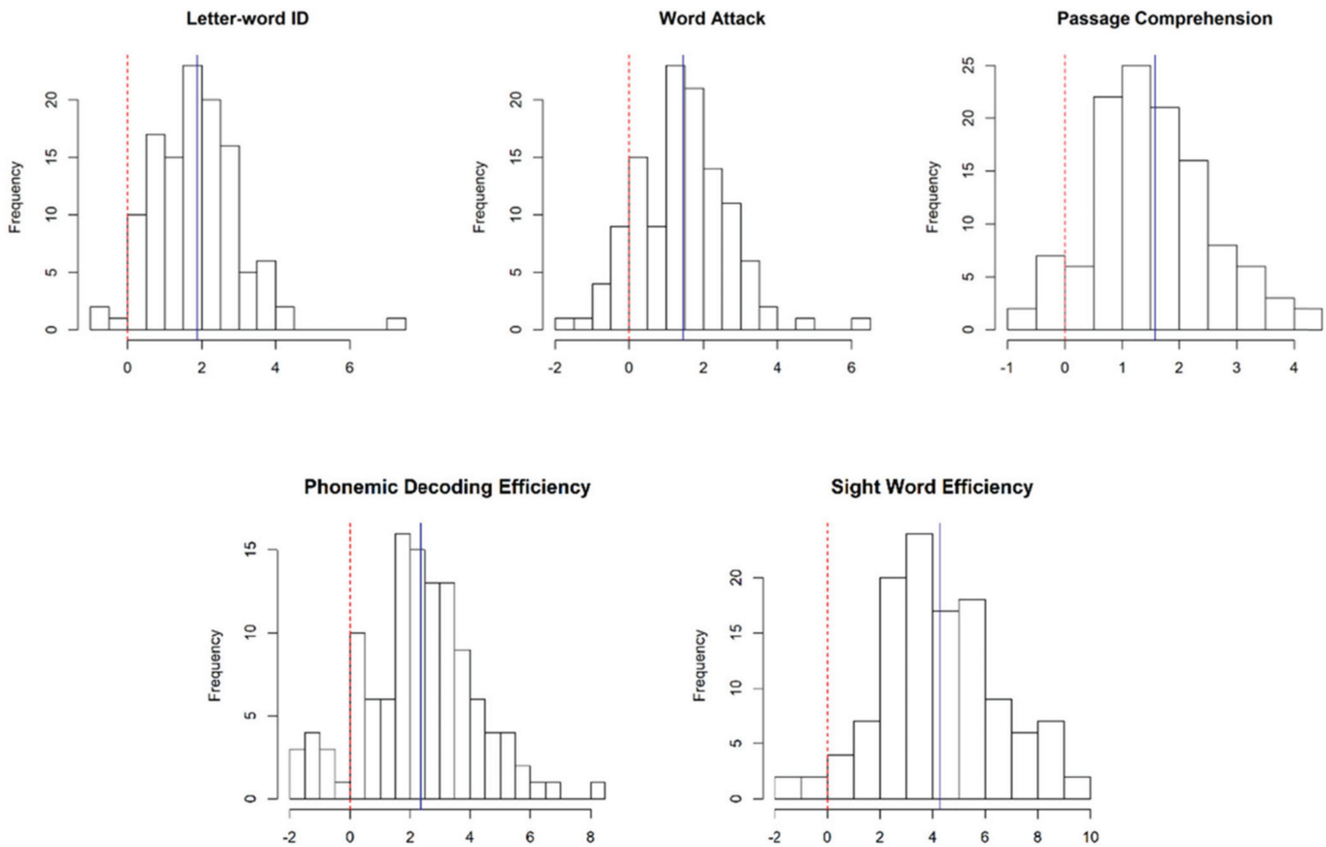
Methodological issues related to the measurement of change are well-documented, not only in studies of response to intervention (Frijters et al., 2013; Fuchs, 2003) but in psychology in general (e.g., Cronbach & Furby, 1970). Here we opted to use a simple operational measure to quantify individual differences in skill gains. For each child, for each of the five outcome measures (three Woodcock-Johnson III and two TOWRE subtests), we ran a simple linear model predicting the subtest raw scores as a function of administration/time-point (0,1,2,3; treated as a numeric variable). Readers can examine the individual-level scores in the different reading subtests as a function of time-point using the interactive plots at https://osf.io/6kwmr/. These models resulted in two metrics for each child in each subtest: an intercept, representing estimated preintervention subtest score; and a slope, representing estimated gain in raw scores as a function of time under intervention (i.e., estimated change in raw score per time-point). Unsurprisingly, the slope scores were generally positive, showing that on average individuals did show gains over the course of the intervention (Figure 1; note that in all five subtests mean slopes were significantly larger than 0, all $ps < .001$). At the same time, as expected given documented individual differences in response to the same intervention used here in previous large-scale studies (e.g., Lovett et al., 2017; Morris et al., 2012), extensive variability was observed in all slope measures. The analyses presented below aim to predict the variability in slope scores, while controlling for variability in the intercept (starting point) values (i.e., predicting residualized gains). Note that although all correlations among the residualized gains in the five reading outcomes were positive and significant, they were small to moderate in size (see Table 1). In the analyses below we therefore examine the predictors of gains in each of these components separately (a decision that was eventually supported by the observed differential predictive effects of O-P and imageability across different reading measures).

Investigating Correlates of Individual Differences in Skill Gains

Here we report the degree to which preintervention measures of reliance on O-P and imageability predict individual-level skill gains and the degree to which these gains in reading over the course of the intervention are associated with changes in reliance on O-P and imageability. This was examined using a series of stepwise regression models where the dependent variable was one of the five slope scores representing gains over the course of intervention in each of the five reading subtests. At a first step, the models included only one predictor—the estimated intercept on the same reading subtest. This was done in order to control for the effect of the preintervention starting point, and to make sure that in the later regression steps we examine predictive values in regards to residualized, rather than raw, change. At a second step, the models also included the independent variables of interest, namely, the predictors from the word reading task: preintervention reliance on O-P regularities and imageability; and the differences between post and preintervention reliance on O-P and imageability (i.e., $\Delta IMG = IMG_{post} - IMG_{pre}$; $\Delta O\text{-}P = O\text{-}P_{post} - O\text{-}P_{pre}$). We examined whether the second-step model improved model fit by comparing the second-step model to the first-step (intercept-only model), using an F test within the *anova* function in R. As a third step, we further examined the possible interactions between changes and starting points in reliance on O-P and imageability: For example, it is possible that children who started with a lower use of either of these dimensions (e.g., minimal O-P), but showed greater gains from pre- to postintervention (e.g., large ΔO-P), are those who had increased reading gains. This interactive model was compared to the additive model (second step), again using an F test. Below we report the better model for each reading subtest, depending on whether adding the interaction terms significantly improved model fit compared to the additive model.[7]

[7] One possible concern in our analytical approach is the relatively high degree of collinearity between predictors (e.g. between starting point and change in both O-P and imageability; and between preintervention O-P and imageability and subtests' intercepts). To make sure the collinearity did not skew the estimates, in the online supplemental materials we run the same analysis here using a pls (rather than ols) regression, which is less susceptible to statistical errors in cases of collinearity. The results were qualitatively similar, thus validating our main findings.

Figure 1
Distribution of Slope Scores (Change in Raw Scores Per Time-Point) in The Three Woodcock-Johnson Subtests (Top) and the Two Towre Subtests (Bottom)



Note. Dashed red lines show the zero point (no change). Solid blue lines depict sample means. See the online article for the color version of this figure.

In what follows we describe the results of the models for each of the five assessment subtests. The results of the models are also summarized in Tables 2, 3, and 4, which present the coefficient estimates of best models for tasks measuring pseudoword reading, word reading, and passage comprehension, respectively. Figures 2, 3, and 4 depict the (raw) correlations between observed significant predictors and residualized gains.

Predicting Gains in Pseudoword Reading

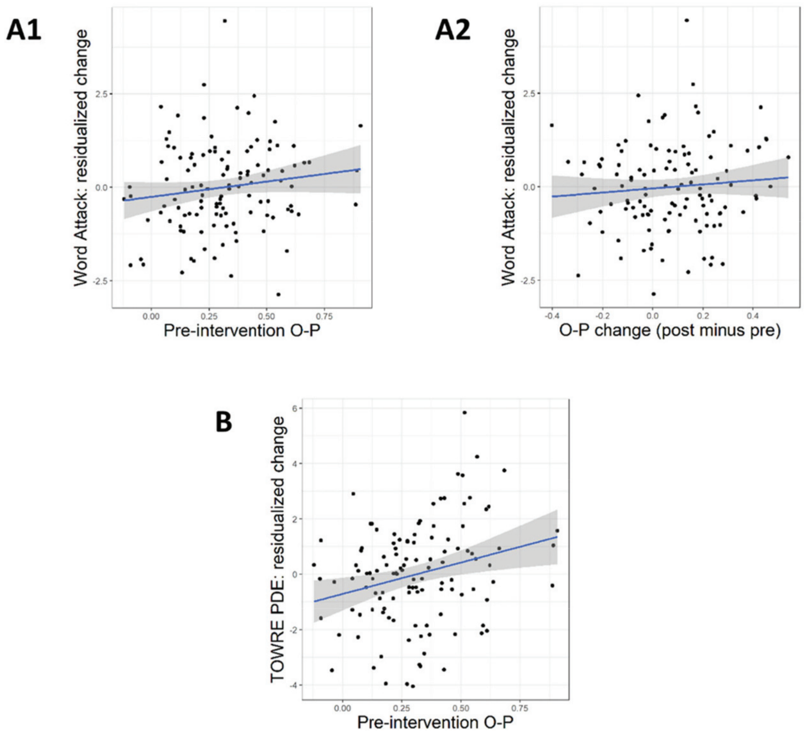
Word Attack (Untimed Test). The first-step model revealed that higher intercepts in Word Attack (starting point) were associated with lesser gains (i.e., slopes) on the same task. Importantly, adding the preintervention and change scores from the word reading task significantly improved model fit, $F(4, 116) = 2.94$, $p = .023$, suggesting that these measures have unique value in predicting residualized intervention gains. Examining the estimated coefficients revealed that stronger reliance on O-P preintervention, and larger increase in O-P over the course of the intervention, were both positively associated with greater skill gain as indexed via Word Attack (see Table 2, and Figure 2 Panels A1 and A2). Adding the interaction terms did not improve model fit in comparison to the additive model ($p = .27$). Note that there was no

Table 1
Correlations Among Residualized Gains (Slopes Controlling for Intercepts) in the Five Reading Measures

Measure	2	3	4	5
1. Woodcock-Johnson: Word Attack	.38	.22	.38	.26
2. Woodcock-Johnson: Letter-Word Identification	—	.34	.39	.37
3. Woodcock-Johnson: Passage Comprehension		—	.19	.29
4. TOWRE: Phonemic Decoding Efficiency			—	.37
5. TOWRE: Sight Word Efficiency				—

Note. All correlations are significant (all $ps < .05$).

Figure 2
Significant Predictors of Skill Gains in Pseudoword Reading



Note. Panels A1 and A2: Predicting residualized gains in Word Attack (untimed test): starting point and change in reliance on O-P. Panel B: Predicting residualized gains in Phonemic Decoding Efficiency (timed test): starting point in reliance on O-P. See the online article for the color version of this figure.

evidence for associations between Word Attack gains and preintervention or change in reliance on imageability.

Phonemic Decoding Efficiency (Timed Test). Here the first step model did not show a significant relation between starting points and gains ($p > .1$), although after adding the measures from the word reading task in the second-step model this (negative) association did reach significance (see Table 2). Adding the measures of reliance on O-P and imageability again improved model fit, $F(4, 116) = 3.61$, $p = .008$. Estimated coefficients showed that similarly to gains in untimed pseudoword naming (i.e., Word Attack), greater preintervention reliance on O-P regularities was associated with greater gains in pseudoword efficiency (Figure 2B). No other significant relations were found, and the interactive model did not have a better fit compared to the additive one ($p = .49$).

Predicting Gains in Word Reading

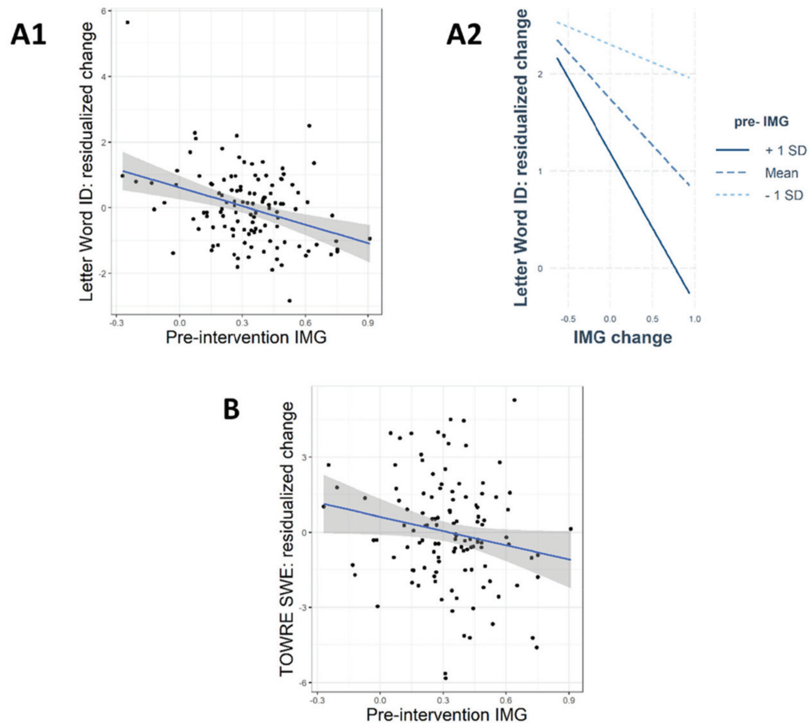
Letter-Word Identification (Untimed Test). Again, higher intercepts (starting point) were associated with lower slopes on the task. Adding the measures from the word reading task again improved model fit, $F(4, 116) = 5.81$, $p < .001$. Moreover, adding the interaction terms further improved the fit in relation to the additive model, $F(2, 110) = 3.89$, $p = .023$. Table 3 includes the coefficients from the interactive model (note that estimated main effects were qualitative similar in the additive and interactive models). As can be seen, less preintervention reliance on imageability was associated with greater gains in Letter-Word Identification (Figure 3, panel A1). In addition, we found a significant interaction between preintervention reliance on imageability and change in the magnitude of the imageability effect over the course of the intervention, such that the impact of a reduction in imageability was more pronounced in children with higher preintervention reliance on imageability (see Figure 3, panel A2). There was also a marginally significant relation ($p = .061$) between a greater decrease in the imageability effect and a greater skill gain. Effects related to reliance on O-P regularities did not reach significance.

Predicting Gains in Sight Word Efficiency (Timed Test). There was a significant (negative) relation between intercept and slope in the first step model ($p < .001$). The second step model again showed improved model fit compared to the intercept-only model, $F(4, 116) = 2.77$, $p = .031$, with greater preintervention reliance on imageability significantly associated with less residualized gains (see Table 3, Figure 3B), in line with the results from the untimed Letter-Word Identification test. All other predictors were not significant, and the interactive model did not improve fit further ($p = .76$).

Predicting Gains in Passage Comprehension

The first-step model showed that higher intercepts were marginally related to lesser slopes ($p = .052$), an effect that reached significance at the second-step model that also includes measures of

Figure 3
Significant Predictors of Skill Gain in Word Reading



Note. Panels A1 and A2: Predicting residualized gains in Letter-Word Identification (untimed test). A1: Starting point in reliance on imageability. A2: The estimated effect of the interaction between preintervention reliance on imageability and change—larger effect of imageability change for individuals who relied more on imageability preintervention. Panel B: Predicting residualized gains in Sight Word Efficiency (timed test): starting point in reliance on imageability. See the online article for the color version of this figure.

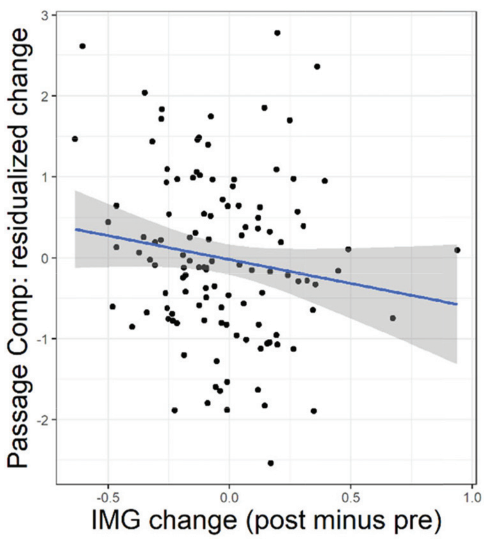
reliance on O-P and imageability (see Table 4). Adding the measures from the word reading task resulted in a marginal improvement in model fit, $F(4, 116) = 2.35$, $p = .058$. Examining the individual predictors showed that greater residualized gains in Passage Comprehension were significantly associated with a greater reduction in reliance on imageability over the course of the intervention (see Figure 4). There was also a marginally significant relation ($p = .058$) between lower preintervention reliance on imageability and greater skill gain in Passage Comprehension. Adding the interaction terms did not improve fit further ($p = .67$).

Controlling for Other Expected Predictors of Intervention Gains

Lastly, we wished to examine the role of other expected predictors of intervention outcomes, and whether our measures from the word reading task account for skill gains beyond these factors. To do so, we added as predictors measures of preintervention levels of vocabulary, auditory attention, rapid automatized naming, and phonological awareness to the models predicting relative gains from preintervention levels (i.e., intercepts) and the measures from the word reading task (i.e., O-P regularities and imageability). Vocabulary was measured using T-scores from the WASI-II (Wechsler, 2011); auditory attention using scaled scores from NEPSY-II (Korkman et al., 2007); rapid automatized naming via standard scores in the colors subtest (Wolf & Denckla, 2005); and phonological awareness using the composite standard score from CTOPP-II (Wagner et al., 2013). For brevity, the full outputs of these models are reported in the online supplemental materials. Overall, the results showed that reliance on O-P regularities and imageability accounted for a significant portion of the variance in intervention gains even when including these additional factors. Thus, we observed qualitatively similar effects of reliance on O-P and imageability as in the original models above (i.e., models without the general preintervention factors; Tables 2, 3, and 4), with all significant effects remaining significant and all insignificant effects but one remaining insignificant. The results also suggest that the general preintervention factors have somewhat limited predictive value compared to the measures from the word reading aloud task, especially in predicting gains in word and pseudoword reading. Concretely, we found that the measures from the word reading aloud task predicted larger unique portions of the variance than the general preintervention factors in all measures of word and pseudoword reading (Letter-Word Identification: $\Delta R^2 = 21.3\%$ vs. 1.4%; Sight Word Efficiency: $\Delta R^2 = 7.8\%$ vs. 6.5%; Word Attack: $\Delta R^2 = 8.7\%$ vs. 6.6%; Phonemic Decoding Efficiency: $\Delta R^2 = 11.4\%$ vs. 3.5%; note that ΔR^2s represent added value when controlling for other

Figure 4

Significant Predictor of Skill Gain in Passage Comprehension: Starting Point in Reliance on Imageability



Note. See the online article for the color version of this figure.

predictors). In predicting gains in Passage Comprehension, the unique portion accounted for by the general predictors was larger (ΔR^2 = 16.9%; attributed mostly to vocabulary scores), but measures from the word reading task still accounted for an additional 7.5% of the variance. Together, these results suggest that our measures of the functional organization of the reading system are unique predictors of relative skill gains, tapping into parts of the variance unaccounted for by "typical" predictors of intervention outcomes.

General Discussion

What are the predictors of success (or failure) of a given reading intervention program for a given individual? This question presents an ongoing challenge for research and practice, with recent studies suggesting that current known predictors only account for a limited portion of the variance in intervention outcomes (e.g., Stuebing et al., 2015). The goal of the current article therefore was to examine a novel set of predictors of intervention gains, comprising proxies of the functional organization of a child's reading system that are theoretically grounded in the triangle model of reading (e.g., Seidenberg & McClelland, 1989). Specifically, given recent behavioral findings relating concurrent reading skills to individuals' reliance on O-P regularities and imageability during an oral word reading task (Siegelman, Rueckl, et al., 2020), we examined whether preintervention profiles and change along these dimensions predict individuals' degree of change during a phonologically-based decoding strategies intervention. Our results show that individuals who preintervention had a profile that is associated with better early reading skills (i.e., more reliance on O-P; less reliance on imageability) were those who generally showed greater gains during the specific type of intervention provided. In parallel, those who showed greater change along these two dimensions in a direction that is associated with better reading skills (i.e., greater increase in O-P; greater decrease in imageability) were generally characterized by greater gains.

From the perspective of the triangle model of reading, these results can be taken to stem from the critical role of sublexical O-P regularities in shaping the organization of the reading system.

Table 2

Regression Models Predicting Individual-Level Gains in Pseudoword Reading Tasks—Word Attack From Woodcock-Johnson (Untimed Test) and Phonemic Decoding Efficiency From TOWRE (Timed)—From Preintervention Reliance on O-P and Imageability and Change in O-P and Imageability

Predictor	β (coefficient)	*SE*	*z* value	*p*-value	ΔR^2
	Dependent variable: Gains in Word Attack (R^2 = 17.5%)				
WA intercept	−0.14	0.03	−4.56	**<.001**	8.8%
Pre-O-P	0.43	0.15	2.80	**.006**	
Pre-IMG	−0.21	0.15	−1.38	.170	8.7%
ΔO-P	0.29	0.13	2.16	**.032**	
ΔIMG	−0.14	0.14	−1.01	.316	
	Dependent variable: Gains in Phonemic Decoding Efficiency (R^2 = 11.4%)				
PDE intercept	−0.07	0.03	−2.23	**.028**	<0.1%
Pre-O-P	0.67	0.25	2.72	**.008**	
Pre-IMG	−0.31	0.23	−1.34	.184	11.4%
ΔO-P	0.02	0.21	0.12	.906	
ΔIMG	−0.05	0.22	−0.24	.809	

Note. When the interaction between change and initial reliance improved model fit, models also include the interaction terms. All models include the estimated intercept (starting point) of the relevant subtest as a control variable. IMG = reliance on imageability; O-P = reliance on O-P regularities (O-P surprisal effect); WA = Word Attack; PDE = Phonemic Decoding Efficiency; Pre- = preintervention values; Δ = change from Pre- to postintervention. Predictors from the word reading task are centered and scaled. R^2 values for the intercept were estimated based on the intercept-only model; R^2 of models with O-P/imageability were calculated by subtracting the R^2 of the full model from that of the intercept-only model. The *p*-values for significant effects ($p < .05$) are shown in bold.

Table 3

Regression Models Predicting Individual-Level Gains in Word Reading Tasks—Letter Word Identification From Woodcock-Johnson (Untimed Test) and Sight Word Efficiency From TOWRE (Timed)—From Preintervention Reliance on O-P and Imageability and Change in O-P and Imageability

Predictor	β (coefficient)	*SE*	*z* value	*p*-value	ΔR^2
Dependent variable: Gains in Letter-Word Identification (R^2 = 27.3%)					
LW intercept	−0.09	0.02	−4.31	**<.001**	6.1%
Pre-O-P	0.10	0.13	0.78	.437	
Pre-IMG	−0.56	0.13	−4.25	**<.001**	
ΔO-P	−0.14	0.12	−1.20	.234	21.3%
ΔIMG	−0.24	0.13	−1.89	.061	
(Pre-O-P) × (ΔO-P)	−0.14	0.09	−1.59	.114	
(Pre-IMG) × (ΔIMG)	−0.18	0.07	−2.51	**.012**	
Dependent variable: Gains in Sight Word Efficiency (R^2 = 21.6%)					
SWE Intercept	−0.09	0.02	−5.06	**<.001**	13.8%
Pre-O-P	0.27	0.26	1.06	.29	
Pre-IMG	−0.54	0.27	−2.05	**.04**	7.8%
ΔO-P	−0.27	0.24	−1.13	.26	
ΔIMG	−0.21	0.25	−0.84	.40	

Note. When the interaction between change and initial reliance improved model fit, models also include the interaction terms. All models include the estimated intercept (starting point) of the relevant subtest as a control variable. IMG = reliance on imageability; O-P = reliance on O-P regularities (O-P surprisal effect); LW = Letter-Word Identification; SWE = Sight Word Efficiency; Pre- = preintervention values; Δ = change from pre- to postintervention. Predictors from the word reading task are centered and scaled. R^2 values for the intercept were estimated based on the intercept-only model; R^2 of models with O-P/imageability were calculated by subtracting the R^2 of the full model from that of the intercept-only model. The *p*-values for significant effects ($p < .05$) are shown in bold.

Given the nature of the statistical structure of alphabetic writing systems, successful assimilation of these regularities is an important step in the early stages of the acquisition process (Harm & Seidenberg, 1999, 2004; Rueckl et al., 2019). Simulation studies show that a lack of an integrity of the phonological system may impact the organization of the reading system by causing an overreliance on whole-word associations ("overfitting") at the expense of poorer attunement to sublexical O-P regularities (Harm et al., 2003; Harm & Seidenberg, 1999; Rueckl et al., 2019). Other simulations suggest that if the O-P system settles on this relatively inefficient organization, the reading system as a whole adopts an organization such that processes that map orthography to phonology via semantics (O-S-P) play a more important role in phonological decoding (Harm & Seidenberg, 2004; Woollams et al., 2007). From this perspective, the magnitude of the imageability effect indexes the contribution of semantics to word naming (Strain et al., 1995), and thus a larger imageability effect for a given child reader is a consequence of a suboptimal organization of the developing reading system.

This interpretation of the imageability effect is important for understanding the directionality of our findings: that is, the fact that reading outcomes at the end of the intervention were positively related to reliance on O-P regularities, but negatively related to reliance on imageability. Note that the directionality of these

Table 4

Regression Models Predicting Individual-Level Gains in Passage Comprehension From Preintervention Reliance on O-P and Imageability and Change in O-P and Imageability

Predictor	β (coefficient)	*SE*	*z* value	*p*-value	ΔR^2
Dependent variable: Gains in Passage Comprehension (R^2 = 10.7%)					
PC intercept	−0.08	0.03	−2.51	**.014**	3.2%
Pre-O-P	0.15	0.13	1.23	.222	
Pre-IMG	−0.25	0.13	−1.92	.058	7.5%
ΔO-P	0.05	0.12	0.41	.684	
ΔIMG	−0.31	0.12	−2.54	**.012**	

Note. When the interaction between change and initial reliance improved model fit, models also include the interaction terms. all models include the estimated intercept (starting point) of the relevant subtest as a control variable. IMG = reliance on imageability; O-P = reliance on O-P regularities (O-P surprisal effect); PC = Passage Comprehension; Pre- = preintervention values; Δ = change from pre- to postintervention. Predictors from the word reading task are centered and scaled. R^2 values for the intercept were estimated based on the intercept-only model; R^2 of models with O-P/imageability were calculated by subtracting the R^2 of the full model from that of the intercept-only model. The *p*-values for significant effects ($p < .05$) are shown in bold.

effects is in line with that observed in the concurrent correlations of reliance on O-P and imageability with reading skills (Siegelman, Rueckl, et al., 2020; and online supplemental materials). Importantly, the negative relations between reading skills/relative gains and imageability, and the positive relations they have with reliance on O-P regularities, are expected from the perspective of the triangle model of reading, where reliance on imageability indexes the contribution of semantics related to a suboptimal organization. A related point is that reliance on imageability should not be conflated with reliance on all types of O-S regularities. Rather, imageability is a nonsystematic (or arbitrary) cue, which does not provide a way to reliably map printed units into meaning. From this perspective, an individual's greater reliance on imageability reflects increased reliance on inefficient O-S processes, which is expected in readers with an underdeveloped system (Siegelman, Rueckl, et al., 2020; see also Pugh et al., 2008; Woollams et al., 2016, for related findings and discussion). This point also raises an intriguing possibility, that reading skills and intervention gains may be positively correlated with alternative measures, not explored here, that tap into individuals' knowledge of the regularities that do exist in the O-S mapping, including in particular morphological relations (see Ulicheva et al., 2020, for preliminary evidence).

Beyond the general direction of the relations between the measures from the word reading task and skill gains, our findings also highlight their differential contributions to predicting intervention outcomes depending on the type of reading measure utilized. Namely, specific intervention outcomes were differentially tied to either the O-P or imageability effect: Greater preintervention and/or increased reliance on O-P were exclusively related to gains in pseudoword reading, while lower preintervention and/or decreased reliance on imageability over the course of the treatment were related to gains in word reading and comprehension (but not to pseudoword reading). At the same time, our own behavioral concurrent (i.e., preintervention) correlational findings were not characterized by such different effects, as both O-P and imageability slope scores were related to all reading subtests (i.e., both O-P and imageability were related to various reading subtests, including word reading, pseudoword reading, and passage comprehension; see Siegelman, Rueckl, et al., 2020; and online supplemental materials). This raises the following question: Why were our measures of reliance on O-P regularities and imageability differentially predictive of gains in different reading measures, whereas concurrently (preintervention) they were related to reading skills across measures?

One possible explanation may be related to nonlinear effects of the organization of the reading system on different outcomes. Thus, it is possible that earlier on in development, some integrity of both the O-P and O-S pathways is necessary to reach basic reading skills (across measures), but that more specific effects relating the two pathways to different components of the outcomes emerge after the reading system is exposed to a larger amount of written input (e.g., as a result of an intense intervention program). This hypothesis can be tested both behaviorally and computationally by examining the relative contribution of reliance on O-P and imageability to different reading outcomes across a longer developmental trajectory.

Another possible explanation is related to the distinction between different types of regularities within the O-P mapping. In the triangle model, the organization of the O-P pathway is shaped by regularities at different grain sizes, from the smaller grapheme–phoneme regularities, to body-rime regularities and lexical regularities at larger grain sizes. According to the overfitting hypothesis discussed above, although factors that reduce the integrity of the phonological system can impair the acquisition of O-P knowledge across grain sizes, they have a greater impact on the learning of smaller regularities (i.e., grapheme–phoneme correspondences). Note that whereas performance on the pseudoword reading assessment depends on knowledge of grapheme–phoneme correspondences, performance on the word reading and comprehension depend on knowledge of multiple types of regularities. Given that our measure of O-P reliance indexes the use of grapheme–phoneme pairings, it may be more closely related to gains in pseudoword naming than word naming or comprehension. In contrast, if the intervention improved a child's knowledge of O-P regularities at the larger grain size, these improvements might not be detected by the measure of O-P reliance but could result in gains on word naming and comprehension (see also Steacy et al., 2020). If greater semantic reliance is a consequence of poor O-P functionality across grain sizes, and if poor O-P functionality limits the potential benefits of a phonologically-weighted intervention, then the observed negative association between semantic reliance and gains on the word reading and comprehension would be expected.

The exact mechanistic explanation behind these differential effects notwithstanding, we believe that they have important methodological implications. Thus, these results highlight the need to separately consider different intervention outcomes, as some individuals may show responses on some outcomes more than on others, reflecting possibly both the developmental progression underlying learning to read and theoretically meaningful variance that is tied to different profiles. In this context, it is interesting to note that the correlations between O-P and imageability were generally limited: O-P and imageability were significantly but moderately negatively correlated preintervention ($r = -.21$, $p = .025$; see online supplemental materials), and the correlation between change scores in the two dimensions was insignificant ($r = .12$, $p = .20$).

To clarify, we do not claim that reliance on O-P versus imageability and different reading outcomes reflect orthogonal components: The reading system is a dynamic, multidimensional system where at each point in development a given cognitive profile is the outcome of complex interactions between its different components. The overfitting hypothesis mentioned above is one example for the reading system's interactive nature, where failure to attune to sublexical O-P relations may lead to greater reliance on whole-word associations (Harm & Seidenberg, 1999), including O-P and O-S associations, which may in turn lead to a stronger influence of semantic properties such as imageability Future studies of intervention outcomes therefore should not only focus on shared gains across measures (e.g., gains in composite scores) but should aim to account for differential variability in gains reflected in different outcomes, while keeping in mind that different component reading skills (e.g., pseudoword and word reading) ultimately impact one another and jointly determine long-term reading outcomes.

A related, broader, open question has to do with the determinants of individual differences in reliance on O-P and imageability: That is, what is the mechanism(s) that gives rise to an individual's organization of the reading system? Theoretically, these individual differences can be tied to differences in readers' quality of phonological and/or orthographic representations (e.g., Harm & Seidenberg, 1999); their oral vocabulary skills (reflecting knowledge of associations between phonology and semantics, which is beneficial for developing an efficient division of labor; Chang & Monaghan, 2019; Chang, Monaghan, et al., 2019); and individual differences in learners' ability to track the statistical relationships between orthographic, phonological, and semantic lexical and sublexical units (i.e., statistical learning computations; see e.g., Sawi & Rueckl, 2019), all of which may be tied to their specific instructional and learning environment experiences and history. In the context of predicting intervention gains, a key finding in our analysis was that the effects of O-P and imageability on skill gains are observed even when controlling for a variety of individual differences measures (which in previous studies were only somewhat linked to intervention responsiveness). Specifically, our data suggest that the role of O-P and imageability in predicting skill gains cannot be attributed simply to individual variability in quality of phonological representations (reflected in measures of phonological awareness), knowledge of phonology–semantics associations (reflected by vocabulary skill; see Chang & Monaghan, 2019), rapid naming, or auditory attention. It is also notable that reliance on O-P and imageability accounted for larger portions of the variance in word and pseudoword reading gains than these common general predictors. Still, more studies are needed to fully unveil the complex relations between all these variables (i.e., how individual differences along these various dimensions give rise to differences in reliance on O-P and imageability), and their relative contribution for intervention outcomes in different populations. Here, too, future behavioral investigations can be aided by computational simulations, which can lesion parts of the reading network in ways that differentially map into these specific mechanisms (e.g., Harm & Seidenberg, 1999). Of particular relevance in this context is computational work that simulates intervention protocols (e.g., Harm et al., 2003), which can be used to examine the effect of a given intervention as a function of network architectures simulating deficits in these different mechanisms.

Before turning to broader implications, it is important to highlight the methodological limitations of the current study. The first limitation is that our study did not aim to have the design characteristics needed to evaluate an aptitude by treatment interaction (i.e., it did not have a control group of children with reading disabilities and random assignment). Hence, our study cannot establish that the effects of variability in O-P or imageability are directly related to response to intervention (i.e., vs. gains over the course of general development or exposure to the educational environment). Future studies could incorporate random assignment and a control group to establish that the relative gain effect was linked specifically to receipt of the intervention. Another methodological limitation has to do with potential measurement error in both outcomes (i.e., gain estimates as a response to the intervention) and predictors (i.e., slope scores from the word reading task). As noted above, issues related to assessing intervention gains are well documented, as measuring change inherently introduces noise (e.g., Cronbach & Furby, 1970). Yet another source of measurement error in our design stems from the slope estimates of reliance on O-P regularities and imageability during the word reading task (even within a time-point). Namely, the reliability of these estimates for a given participant most likely vary as a function of their mean accuracy on the task and may be limited among children at the edges of the accuracy distribution (i.e., with very high or low performance). Together, these two sources of measurement error most likely led to an attenuation of the observed associations between our predictors and the intervention outcomes. Indeed, although the portion of variance explained by our measures does not fall short (and if anything is higher in most cases) than the estimated effect sizes of other individual-level factors (e.g., Stuebing et al., 2015), it is obvious that much variability is still left unaccounted for (see scatter plots in Figures 2 to 4). Future research is challenged with improving the reliability of estimates of both predictors and outcomes, by improving the word reading task (e.g., by using an adaptive design where mean accuracy is matched across subjects), and by adopting more sophisticated statistical models for estimating change and its relation to individual-level factors (e.g., Frijters et al., 2013). In regards to the latter, our findings highlight the importance of utilizing approaches that allow for the detection of effects that vary across different intervention outcomes.

Lastly, we wish to discuss the implications of the findings for educational practice, where the main goal of studies such as ours is to identify the relations between an individual's attributes and an optimal remediation tailored for their specific needs. We believe that our study takes a step toward this goal by tying the component pathways of a child's reading system to their response to a given phonologically-weighted intervention. Thus, our study sheds new light on why some struggling readers respond better than others to the same intervention. It also provides a potentially useful strategy to tap into the components of the reading system using a simple word reading aloud task, which may be used in typical educational settings (pending modifications to the task before it can be used to capture variation among individuals with diverse reading skills, as mentioned above). Yet a critical question that is left unanswered is what would be an effective intervention to those who show limited gains to the currently investigated program, or how would they have responded to a differently focused intervention. In other words, what would be an optimal intervention program for readers with a profile that is correlated here with lesser gains (e.g., no preintervention reliance on O-P and increased reliance on imageability)? One option is that these readers simply need a more intense intervention of the same type (increased dose; e.g., more exposure to O-P until they assimilate the regularities in this mapping). Alternatively, these readers may require a remediation program that includes a larger emphasis on additional intervention components (e.g., one that, in addition to sublexical phonology, includes a larger focus on morphology, or on decreasing reliance on imageability, however that may be done). Future work is left with examining the links between response to intervention to different programs and the individual's functional organization of the reading system, with the goal of eventually a priori identifying the optimal intervention program for each individual based on their preintervention cognitive profile.

References

Al Otaiba, S., & Fuchs, D. (2002). Characteristics of children who are unresponsive to early literacy intervention: A review of the literature. *Remedial and Special Education*, *23*(5), 300–316. https://doi.org/10.1177/07419325020230050501

Arrington, C. N., Malins, J. G., Winter, R., Mencl, W. E., Pugh, K. R., & Morris, R. (2019). Examining individual differences in reading and attentional control networks utilizing an oddball fMRI task. *Developmental Cognitive Neuroscience*, *38*, 100674. https://doi.org/10.1016/j.dcn.2019.100674

Castles, A., Rastle, K., & Nation, K. (2018). Ending the reading wars: Reading acquisition from novice to expert. *Psychological Science in the Public Interest*, *19*(1), 5–51. https://doi.org/10.1177/1529100618772271

Chang, Y. N., & Monaghan, P. (2019). Quantity and diversity of preliteracy language exposure both affect literacy development: Evidence from a computational model of reading. *Scientific Studies of Reading*, *23*(3), 235–253. https://doi.org/10.1080/10888438.2018.1529177

Chang, Y. N., Monaghan, P., & Welbourne, S. (2019). A computational model of reading across development: Effects of literacy onset on language processing. *Journal of Memory and Language*, *108*, Article 104025. https://doi.org/10.1016/j.jml.2019.05.003

Coltheart, V., Laxon, V. J., & Keating, C. (1988). Effects of word imageability and age of acquisition on children's reading. *British Journal of Psychology*, *79*(1), 1–12. https://doi.org/10.1111/j.2044-8295.1988.tb02270.x

Cronbach, L. J., & Furby, L. (1970). How we should measure "change": Or should we? *Psychological Bulletin*, *74*(1), 68–80. https://doi.org/10.1037/h0029382

Davies, R. A. I., Birchenough, J. M. H., Arnell, R., Grimmond, D., & Houlson, S. (2017). Reading through the life span: Individual differences in psycholinguistic effects. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *43*(8), 1298–1338. https://doi.org/10.1037/xlm0000366

Duff, F. J., & Clarke, P. J. (2011). Practitioner review: Reading disorders: What are the effective interventions and how should they be implemented and evaluated? *Journal of Child Psychology and Psychiatry*, *52*(1), 3–12. https://doi.org/10.1111/j.1469-7610.2010.02310.x

Frijters, J. C., Lovett, M. W., Sevcik, R. A., & Morris, R. D. (2013). Four methods of identifying change in the context of a multiple component reading intervention for struggling middle school readers. *Reading and Writing*, *26*(4), 539–563. https://doi.org/10.1007/s11145-012-9418-z

Frijters, J. C., Lovett, M. W., Steinbach, K. A., Wolf, M., Sevcik, R. A., & Morris, R. D. (2011). Neurocognitive predictors of reading outcomes for children with reading disabilities. *Journal of Learning Disabilities*, *44*(2), 150–166. https://doi.org/10.1177/0022219410391185

Fuchs, L. S. (2003). Assessing intervention responsiveness: Conceptual and technical issues. *Learning Disabilities Research & Practice*, *18*(3), 172–186. https://doi.org/10.1111/1540-5826.00073

Galuschka, K., Ise, E., Krick, K., & Schulte-Körne, G. (2014). Effectiveness of treatment approaches for children and adolescents with reading disabilities: A meta-analysis of randomized controlled trials. *PLoS ONE*, *9*(2), Article e105843. https://doi.org/10.1371/journal.pone.0089900

Glushko, R. J. (1979). The organization and activation of orthographic knowledge in reading aloud. *Journal of Experimental Psychology: Human Perception and Performance*, *5*(4), 674–691. https://doi.org/10.1037/0096-1523.5.4.674

Harm, M. W., & Seidenberg, M. S. (1999). Phonology, reading acquisition, and dyslexia: Insights from connectionist models. *Psychological Review*, *106*(3), 491–528. https://doi.org/10.1037/0033-295X.106.3.491

Harm, M. W., & Seidenberg, M. S. (2004). Computing the meanings of words in reading: Cooperative division of labor between visual and phonological processes. *Psychological Review*, *111*(3), 662–720. https://doi.org/10.1037/0033-295X.111.3.662

Harm, M. W., McCandliss, B. D., & Seidenberg, M. S. (2003). Modeling the successes and failures of interventions for disabled readers. *Scientific Studies of Reading*, *7*(2), 155–182. https://doi.org/10.1207/s1532799xssr0702_3

Jared, D., McRae, K., & Seidenberg, M. S. (1990). The basis of consistency effects in word naming. *Journal of Memory and Language*, *29*(6), 687–715. https://doi.org/10.1016/0749-596X(90)90044-Z

Klose, A. E., Schwartz, S., & Brown, J. W. M. (1983). The imageability effect in good and poor readers. *Bulletin of the Psychonomic Society*, *21*(6), 446–448. https://doi.org/10.3758/BF03330004

Korkman, M., Kirk, U., & Kemp, S. (2007). *NEPSY—Second Edition (Nepsy-II)*. Harcourt Assessment.

Lovett, M. W., Borden, S. L., Lacerenza, L., Frijters, J. C., Steinbach, K. A., & De Palma, M. (2000). Components of effective remediation for developmental reading disabilities: Combining phonological and strategy-based instruction to improve outcomes. *Journal of Educational Psychology*, *92*(2), 263–283. https://doi.org/10.1037/0022-0663.92.2.263

Lovett, M. W., Frijters, J. C., Wolf, M., Steinbach, K. A., Sevcik, R. A., & Morris, R. D. (2017). Early intervention for children at risk for reading disabilities: The impact of grade at intervention and individual differences on intervention outcomes. *Journal of Educational Psychology*, *44*(2), 150–166. https://doi.org/10.1037/edu0000181

Lovett, M. W., Lacerenza, L., & Borden, S. L. (2000). Putting struggling readers on the PHAST track: A program to integrate phonological and strategy-based remedial reading instruction and maximize outcomes. *Journal of Learning Disabilities*, *33*(5), 458–476. https://doi.org/10.1177/002221940003300507

Lovett, M. W., Lacerenza, L., Murphy, D., Steinbach, K. A., De Palma, M., & Frijters, J. C. (2005). Importance of multiple-component interventions for children and adolescents who are struggling readers. In J. Gilger, and S. Richardson (Eds.), *Research-based education and intervention: What we need to know* (pp. 67–102). International Dyslexia Association.

McArthur, G., Eve, P. M., Jones, K., Banales, E., Kohnen, S., Anandakumar, T., Larsen, L., Marinus, E., Wang, H.-C., & Castles, A. (2012). Phonics training for English-speaking poor readers. *Cochrane Database of Systematic Reviews*, *12*, Article CD009115. https://doi.org/10.1002/14651858.cd009115.pub2

Morris, R. D., Lovett, M. W., Wolf, M., Sevcik, R. A., Steinbach, K. A., Frijters, J. C., & Shapiro, M. B. (2012). Multiple-component remediation for developmental reading disabilities: IQ, socioeconomic status, and race as factors in remedial outcome. *Journal of Learning Disabilities*, *45*(2), 99–127. https://doi.org/10.1177/0022219409355472

National Reading Panel. (2000). *Teaching children to read: An evidence-based assessment of the scientific research literature on reading and its implications for reading instruction* (NIH Publication No. 00-4769). U.S. Government Printing Office. https://doi.org/10.1002/ppul.1950070418

National Research Council. (1998). *Preventing reading difficulties in young children*. The National Acadenics Press. https://doi.org/10.17226/6023

Paivio, A., Yuille, J. C., & Madigan, S. A. (1968). Concreteness, imagery, and meaningfulness values for 925 nouns. *Journal of Experimental Psychology*, *76*(1, Pt. 2), 1–25. https://doi.org/10.1037/h0025327

Plaut, D. C., McClelland, J. L., Seidenberg, M. S., & Patterson, K. (1996). Understanding normal and impaired word reading: Computational principles in quasi-regular domains. *Psychological Review*, *103*(1), 56–115. https://doi.org/10.1037/0033-295X.103.1.56

Pugh, K. R., Frost, S. J., Sandak, R., Landi, N., Rueckl, J. G., Constable, R. T., Seidenberg, M. S., Fulbright, R. K., Katz, L., & Mencl, W. E. (2008). Effects of stimulus difficulty and repetition on printed word identification: An fMRI comparison of nonimpaired and reading-disabled adolescent cohorts. *Journal of Cognitive Neuroscience*, *20*(7), 1146–1160. https://doi.org/10.1162/jocn.2008.20079

Rueckl, J. G. (2016). Toward a theory of variation in the organization of the word reading system. *Scientific Studies of Reading*, *20*(1), 86–97. https://doi.org/10.1080/10888438.2015.1103741

Rueckl, J. G., Zevin, J. D., & Wolf VII, H. (2019). Using computational techniques to model and better understand developmental word-reading disorders (i.e., dyslexia). In J. Washington, D. Compton, & P. McCardle (Eds.), *Dyslexia: Revisiting etiology, diagnosis, treatment, and policy* (pp. 57–70). Brookes.

Sawi, O. M., & Rueckl, J. (2019). Reading and the neurocognitive bases of statistical learning. *Scientific Studies of Reading*, *23*(1), 8–23. https://doi.org/10.1080/10888438.2018.1457681

Scheltinga, F., van der Leij, A., & Struiksma, C. (2010). Predictors of response to intervention of word reading fluency in Dutch. *Journal of Learning Disabilities*, *43*(3), 212–228. https://doi.org/10.1177/0022219409345015

Seidenberg, M. S. (1985). The time course of phonological code activation in two writing systems. *Cognition*, *19*(1), 1–30. https://doi.org/10.1016/0010-0277(85)90029-0

Seidenberg, M. S., & McClelland, J. L. (1989). A distributed, developmental model of word recognition and naming. *Psychological Review*, *96*(4), 523–568. https://doi.org/10.1037/0033-295X.96.4.523

Siegelman, N., Kearns, D. M., & Rueckl, J. G. (2020). Using information-theoretic measures to characterize the structure of the writing system: The case of orthographic–phonological regularities in English. *Behavior Research Methods*, *52*(3), 1292–1312. https://doi.org/10.3758/s13428-019-01317-y

Siegelman, N., Rueckl, J. G., Steacy, L. M., Frost, S. J., Van den Bunt, M., Zevin, J. D., Seidenberg, M. S., Pugh, K. R., Compton, D. L., & Morris, R. D. (2020). Individual differences in learning the regularities between orthography, phonology and semantics predict early reading skills. *Journal of Memory and Language*, *114*, Article 104145. https://doi.org/10.1016/j.jml.2020.104145

Steacy, L. M., Edwards, A. A., Rueckl, J. G., Petscher, Y., & Compton, D. L. (2020). Modeling and visualizing the codevelopment of word and nonword reading in children from first through fourth grade: Informing developmental trajectories of children with dyslexia. *Child Development*. Advance online publication. https://doi.org/10.1111/cdev.13468

Strain, E., & Herdman, C. M. (1999). Imageability effects in word naming: An individual differences analysis. *Canadian Journal of Experimental Psychology*, *53*(4), 347–359. https://doi.org/10.1037/h0087322

Strain, E., Patterson, K., & Seidenberg, M. S. (1995). Semantic effects in single-word naming. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *21*(5), 1140–1154. https://doi.org/10.1037/0278-7393.21.5.1140

Stuebing, K. K., Barth, A. E., Trahan, L. H., Reddy, R. R., Miciak, J., & Fletcher, J. M. (2015). Are child cognitive characteristics strong predictors of responses to intervention? A meta-analysis. *Review of Educational Research*, *85*(3), 395–429. https://doi.org/10.3102/0034654314555996

Torgesen, J. K. (2000). Individual differences in response to early interventions in reading: The lingering problem of treatment resisters. *Learning Disabilities Research & Practice*, *15*(1), 55–64. https://doi.org/10.1207/sldrp1501_6

Torgesen, J. K., Wagner, R., & Rashotte, C. (2012). *TOWRE 2: Test of word reading efficieny*. Pro-Ed Inc.

Torgesen, J. K., Wagner, R. K., Rashotte, C. A., Lindamood, P., Rose, E., Conway, T., & Garvan, C. (1999). Preventing reading failure in young children with phonological processing disabilities: Group and individual responses to instruction. *Journal of Educational Psychology*, *91*(4), 579–593. https://doi.org/10.1037/0022-0663.91.4.579

Tran, L., Sanchez, T., Arellano, B., & Swanson, H. L. (2011). A meta-analysis of the RTI literature for children at risk for reading disabilities. *Journal of Learning Disabilities*, *44*(3), 283–295. https://doi.org/10.1177/0022219410378447

Ulicheva, A., Marelli, M., & Rastle, K. (2020). Sensitivity to meaningful regularities acquired through experience. *Morphology*. Advance online publication. https://doi.org/10.1007/s11525-020-09363-5

van der Kleij, S. W., Groen, M. A., Segers, E., & Verhoeven, L. (2019). Sequential implicit learning ability predicts growth in reading skills in typical readers and children with dyslexia. *Scientific Studies of Reading*, *23*(1), 77–88. https://doi.org/10.1080/10888438.2018.1491582

Vaughn, S., Capin, P., Scammacca, N., Roberts, G., Cirino, P., & Fletcher, J. M. (2020). the critical role of word reading as a predictor of response to intervention. *Journal of Learning Disabilities*, *53*(6), 415–427. https://doi.org/10.1177/0022219419891412

Wagner, R. K., Torgesen, J. K., Rashotte, C. A., & Pearson, N. A. (2013). *Comprehensive test of phonological processing* (2nd ed.). Pro-Ed Inc.

Wanzek, J., Vaughn, S., Scammacca, N., Gatlin, B., Walker, M. A., & Capin, P. (2016). Meta-analyses of the effects of Tier 2 type reading interventions in Grades K–3. *Educational Psychology Review*, *28*(3), 551–576. https://doi.org/10.1007/s10648-015-9321-7

Wanzek, J., Vaughn, S., Scammacca, N. K., Metz, K., Murray, C. S., Roberts, G., & Danielson, L. (2013). Extensive reading interventions for students with reading difficulties after Grade 3. *Review of Educational Research*, *83*(2), 163–195. https://doi.org/10.3102/0034654313477212

Wechsler, D. (2011). *Wechsler Abbreviated Scale of Intelligence, Second Edition (WASI-II)*. NCS Pearson.

Wolf, M., & Denckla, M. B. (2005). *Rapid automatized naming and rapid alternating stimulus tests*. Pro-Ed Inc.

Woodcock, R. W., McGrew, K. S., & Mather, N. (2001). *Woodcock-Johnson III Tests of Achievement*. Riverside.

Woollams, A. M., Lambon Ralph, M. A., Madrid, G., & Patterson, K. E. (2016). Do you read how I read? Systematic individual differences in semantic reliance amongst normal readers. *Frontiers in Psychology*, *7*, Article 1757. https://doi.org/10.3389/fpsyg.2016.01757

Woollams, A. M., Ralph, M. A. L., Plaut, D. C., & Patterson, K. (2007). SD-squared: On the association between semantic dementia and surface dyslexia. *Psychological Review*, *114*(2), 316–339. https://doi.org/10.1037/0033-295X.114.2.316

Zeno, S., Ivens, S. H., Millard, R. T., & Duvvuri, R. (1995). *The educator's word frequency guide*. Touchstone Applied Science Associates.

Zevin, J. D., & Seidenberg, M. S. (2006). Simulating consistency effects and individual differences in nonword naming: A comparison of current models. *Journal of Memory and Language*, *54*(2), 145–160. https://doi.org/10.1016/j.jml.2005.08.002

Received August 4, 2020
Revision received February 3, 2021
Accepted February 6, 2021 ■