



Published in final edited form as:

J Abnorm Psychol. 2012 February ; 121(1): 212–224. doi:10.1037/a0025823.

Individual Prediction of Dyslexia by Single vs. Multiple Deficit Models

Bruce F. Pennington,
University of Denver

Laura Santerre-Lemmon,
University of Denver

Jennifer Rosenberg,
University of Denver

Beatriz MacDonald,
University of Denver

Richard Boda,
University of Colorado Denver

Angela Friend,
University of Colorado Denver

Daniel Leopold,
University of Denver

Stefan Samuelsson,
Linköping University and Stavanger University

Brian Byrne,
University of New England and Linköping University

Erik G. Willcutt, and
University of Colorado Boulder

Richard K. Olson
University of Colorado Boulder and Linköping University

Abstract

The overall goals of this study were to test single vs. multiple cognitive deficit models of dyslexia (reading disability) at the level of individual cases and to determine the clinical utility of these models for prediction and diagnosis of dyslexia. To accomplish these goals, we tested five cognitive models of dyslexia: two single-deficit models, two multiple-deficit models, and one hybrid model in two large population-based samples, one cross-sectional (Colorado Learning Disability Research Center—CLDRC) and one longitudinal (International longitudinal Twin Study—ILTS). The cognitive deficits included in these cognitive models were in phonological awareness, language skill, and processing speed and/ or naming speed. To determine whether an individual case fit one of these models, we used two methods: 1) the presence or absence of the predicted cognitive deficits, and 2) whether the individual's level of reading skill best fit the regression equation with the relevant cognitive predictors (i.e. whether their reading skill was proportional to those cognitive predictors.) We found that roughly equal proportions of cases met both tests of model fit for the multiple deficit models (30–36%) and single deficit models (24–

28%); hence, the hybrid model provided the best overall fit to the data. The remaining roughly 40% of cases in each sample lacked the deficit or deficits that corresponded with their best fitting regression model. We discuss the clinical implications of these results for both diagnosis of school age children and preschool prediction of children at risk for dyslexia.

Of all the behaviorally-defined disorders that affect children, we probably have the best theoretical understanding of dyslexia. Dyslexia is defined as a neurodevelopmental disorder characterized by a deficit in the accurate or fluent decoding of single printed words that is not accounted for by a specific sensory deficit or more general intellectual impairment (IDA, 2002). This paper examines two issues: 1) How should we use individual patient data to test theoretical models of dyslexia? and 2) How should research on dyslexia affect the clinician's decision making process with an individual child?

At a theoretical level, dyslexia researchers have posited both single (e.g. Ramus et al., 2003) and multiple deficit (e.g. Bishop & Snowling, 2004; Pennington, 2006) models of dyslexia. What is less known, however, is the extent to which the predictions made by these models using group data can be applied to individual cases. Specific cases can be “acid” tests for the validity of theoretical models of a disorder, especially models that make strong predictions. For example, a single deficit model of dyslexia that posits that a deficit in phonemic awareness (PA) is necessary and sufficient to cause dyslexia will be significantly challenged if substantial numbers of individual cases of dyslexia are found that lack a deficit in PA. Such cases can lead to a reconceptualization or expansion of a theoretical model, or even to a change in the nosology of a disorder (e.g., proposing distinct subtypes or narrowing defining characteristics).

In terms of clinical practice, this study addressed these questions: 1) How successful will a diagnostician be who uses these theoretical models in deciding which individuals will have or currently have dyslexia? 2) In the case of current dyslexia, should a clinician only use reading scores or is the diagnosis strengthened by the fit of the child to a theoretical model? The fact that cognitive models of dyslexia at the group level have been so successful may have made clinicians overconfident about how well they understand each individual case of dyslexia. In formulating an individual case, there may be an inevitable tendency to highlight the evidence which supports the diagnostician's implicit or explicit theoretical model of the disorder and to explain away evidence that does not, which is an example of the availability heuristic in clinical diagnosis (Groopman, 2007). To address these theoretical and applied questions, the current study employed a multiple case study approach, like that used by Ramus et al. (2003) to test the applicability of various current theoretical models of dyslexia (including a phonological deficit model), but with many more cases and across two large samples.

Case study vs. group level approach to dyslexia research

A case study approach is the complement of a multiple regression or structural equation modeling (SEM) approach at the group level, which tests which combination of cognitive predictors accounts for the most variance in reading skill in a population. Using an SEM approach to model the relation between dyslexia and ADHD symptom dimensions in a population, McGrath et al. (2011) showed that three cognitive latent traits – phoneme awareness, processing speed, and naming speed – accounted for 75% of the variance in a latent reading trait, and was a significantly better fit than models with only one or two cognitive predictors. Processing speed made a stronger contribution than naming speed and predicted both timed and untimed reading, whereas naming speed was only significant in predicting timed reading. McGrath et al. (2011) also found the familiar result that phoneme awareness (PA) made the biggest contribution to predicting reading skill. While McGrath et

al.'s (2011) multiple predictor model of reading skill indicates that PA is not *sufficient* to explain reading skill in the population, these results do not tell us if a deficit in PA is *necessary* or sufficient to explain individual cases of dyslexia.

Hence, examining individual cases is an important complementary approach to group level analyses, because even though the fit of a multiple predictor model to the population variance in reading may be high, such a result does not necessarily tell us about what is happening at the level of individual dyslexic cases. Clinicians need to know whether the overall fit of a multiple predictor model is good because nearly every individual fits the same multiple predictor model or because subsets of individuals fit different sub-models, all of which are encompassed in the larger multiple predictor model. For example, it could be the case that some individuals' reading skill may be adequately explained by a specific single predictor, while other individuals' reading skill is explained by a different single predictor. There may also be additional individuals that require multiple predictors to explain their reading performance. As long as all the relevant predictors are incorporated into the SEM model, the overall group level fit to the data will be maximized. However, in this example, the group level results would be misleading, since it would mask the presence of subgroups of individuals, some of whom do not require a multiple deficit model to explain their particular level of reading skill. These different patterns of model fit across individuals, which can only be gleaned when examining individual cases, could potentially define valid subtypes of a disorder (or not). In addition, there may be individuals (dyslexic or not) who are not explained by the SEM model at all. These individuals, as mentioned previously, can provide an acid test for a theoretical model of dyslexia, since the latter is assumed to explain virtually all cases, not just a majority.

Theoretical models of dyslexia under scrutiny in this investigation

This study addresses the theoretically important question of whether single vs. multiple deficit models (or a combination) best account for individual cases of dyslexia. As reviewed elsewhere (e.g. Bishop & Snowling, 2004; Pennington & Bishop, 2009), there are multiple competing models of dyslexia, some with single and some with multiple cognitive predictors. The models tested here all involve the cognitive predictors of reading skill that have been best supported by previous research (e.g. Scarborough, 1998). These include phonemic awareness, language skills (both semantic and syntactic/morphological), naming speed, and processing speed.

Because this study is about dyslexia, it used measures of single word accuracy and fluency, and the results are not directly relevant to children who have poor reading comprehension. Because reading comprehension depends on single word reading skill (Gough & Tunmer, 1986), many children with dyslexia will also perform poorly on reading comprehension.

It should be noted that the issue of single vs. multiple deficit models is not resolved in the dyslexia literature, with phonological core deficit and double deficit accounts still prevalent (e.g., Wolff & Bower, 1999; Ramus, 2003), in addition to the multiple deficit models proposed more recently. So, tests of single vs. multiple deficit models of dyslexia at both the group and individual level is important for current theoretical controversies in dyslexia research and for clinical practice. For instance, the automaticity deficit in dyslexia is not well explained by existing theories. Is this deficit a result of slow processing speed, slow learning of new mappings between phonology and orthography, or poor general language skills? A second key issue is whether dyslexia is exclusively a linguistic disorder or whether non-linguistic cognitive risk factors contribute to it see (Peterson & Pennington, in review), for a discussion of these theoretical issues). With regard to this second issue, we recently found that a phoneme awareness deficit is not sufficient to produce dyslexia in children

identified at preschool age with speech sound disorder - SSD (Peterson, et.al., 2009). Instead, a second deficit in language skill accounted for whether SSD children developed dyslexia. This finding seriously challenges the single phonological deficit model of dyslexia and has implications for both preschool screening and later diagnosis of dyslexia.

In what follows, competing single and multiple deficit models of dyslexia are explained, and predictions arising from each with regard to the case study approach are elaborated. These models are presented roughly in order of parsimony from the most to the least constrained. Thus, the first model will be the easiest model to falsify and the last will be the hardest to falsify.

These models differ on two key issues:

1. Whether a single deficit is necessary and sufficient to cause dyslexia or not;
2. Whether a deficit in PA is necessary to cause dyslexia or not.

Crossing these two dimensions produces four models (See Table 1). The fifth, hybrid model encompasses all four possibilities as different pathways to dyslexia.

As will be explained in more detail later, these theoretical models were tested with two methods 1) presence of the predicted cognitive deficit or deficits in individual cases, 2) fit of individual cases to the predicted regression equation with either single or multiple cognitive predictors. So, the two single deficit models predicted both the presence of a single cognitive deficit (e.g. in PA for Model 1 or for different single cognitive deficits for different subgroups of dyslexics for Model 2) and that individual dyslexics with the predicted cognitive deficit would also best fit the corresponding single cognitive predictor regression equation. In contrast, multiple deficit models predicted the presence of multiple cognitive deficits and fit to the corresponding multiple cognitive predictor equation. The specific cognitive deficits and regression fits predicted by each of the four theoretical models is explained next.

Single Deficit Models

1) Single phonological deficit: A deficit in phoneme awareness (PA) is necessary and sufficient to cause dyslexia.

Predictions

- A) PA as a single predictor in a regression equation will maximally predict individual differences in reading skill and other predictors will lack incremental validity beyond PA.
- B) A PA deficit will maximally predict dyslexic vs. non-dyslexic cases and other predictors will not add to that prediction.
- C) The large majority of dyslexic cases will have a PA deficit and will best fit a PA regression equation, and the remaining cases will not fit another single or multiple predictor regression equation.

2) Single deficit subtypes: Other deficits besides PA, such as deficits in processing speed (PS), naming speed (NS), or language skill (L), are sufficient to cause dyslexia.

Predictions

- A) Other predictors besides PA will have substantial incremental validity in predicting individual differences in reading skill in a multiple regression.

- B) But, the large majority of dyslexic cases will have a single deficit in either PA, PS/NS, or L and will best fit that particular single predictor model. The remaining cases will not fit a multiple deficit model.

Multiple Deficit Models

3) Phonological core, multiple deficit, multiple predictor model: A single PA deficit is necessary but not sufficient to produce dyslexia. So, there must be at least two deficits, one of which is in PA.

Predictions

- A) PA will be the strongest predictor of individual differences in reading skill, but other predictors will have some incremental validity.
- B) The large majority of dyslexic cases will have a PA deficit and at least one other deficit and will fit the multiple predictor model in A). The remaining cases will not fit a single deficit model.

4) Multiple deficit, multiple predictor model: A single deficit is not sufficient to cause dyslexia; at least two deficits are needed.

Predictions

- A) Other predictors besides PA will have substantial incremental validity in predicting individual differences in reading skill in a multiple regression (same as 2A).
- B) The large majority of dyslexic cases will have multiple deficits and will best fit a multiple predictor model. The remaining cases will not fit a single deficit model.

5) Hybrid model (subgroups of individuals with dyslexia will fit each of the four models above.)

There are multiple possible pathways to dyslexia, some involving single deficits and some involving multiple deficits.

Predictions

- A) Other predictors besides PA will have substantial incremental validity in predicting reading skill in a multiple regression (same as 2A and 4A).
- B) Substantial numbers of dyslexic individuals will fit single deficit models, but substantial numbers will fit multiple deficit models.

As can be seen, these models are partially nested. Hence, Model 1 is encompassed by Model 2, except that it is restricted to a subset of dyslexics in Model 2. Similarly, Model 3 is encompassed by Model 4, except that it is restricted to a subset of dyslexics in Model 4. Finally, Models 2 and 4 are encompassed by Model 5, except that they are restricted to subsets of dyslexia.

We will test these models at the level of individual cases in two large samples. There are a number of important reasons for using the two large samples in this study. First, one sample is cross sectional, while the other is longitudinal. The former has a larger age range, allowing us to understand whether predictors of reading change across development, while the latter can be used to understand the predictive power of cognitive variables measured in kindergarten for later reading skill as the outcome. Second, since the two samples used similar, but not identical, measures of the various constructs of interest, comparing the results across samples allows us see if the findings are robust and not just due to method

variance. Third, by having two large, population-based samples, we increase the confidence and generalizability of the results, reducing the possibility that we have not adequately canvassed the entire landscape of individual cases with dyslexia. Although both samples are comprised of twins (selected at random from a twin pair), no premature twins were included in the analyses. Furthermore, previous studies have shown that the effect of twinning is quite small in terms of its impact on academic performance. Specifically, Christensen et al., (2006) showed that academic performance in twin cohorts is similar to that of singletons, and that birth weight had a minimal effect on academic performance. Thus, the results of this study should extend to the non-twin population.

Methods

Statistical issues when using a case study approach

In using a case study approach, we need to consider statistical limits on the possible fit of individual cases to models. These limits derive from 1) the reliabilities of the predictor and outcome measures; 2) the correlations between predictors and outcomes; and 3) the imposition of categorical cutoffs. Unless all reliabilities are perfect (i.e. 1.0) and the multiple correlation between predictors and outcome is also perfect, there will inevitably be false positives and false negatives in predicting individual cases.

The analytic strategy that follows allows us to see which model has the best fit and how close the best fitting model comes to the theoretical maximum dictated by the reliabilities of the measures and the correlations among constructs.

Samples

The participants in this study were randomly selected individuals (only one per twin pair) drawn from two large twin studies: 1) the cross-sectional Colorado Learning Disability Research Center (CLDRC) study of school-age twins (DeFries *et al.*, 1997) over-selected for dyslexia and ADHD, and 2) the longitudinal International Longitudinal Twin Study (ILTS), which is comprised of population samples of preschool age twins from three countries (Australia, Norway, and USA) followed into early school age (Byrne *et al.*, 2002). In the ILTS, we picked all pairs who had the necessary cognitive and reading data at the relevant two time points (end of Kindergarten for deficit factors and end of First grade for reading). There were 809 such twin pairs, and we randomly selected one twin from each pair for a study sample of 809 individuals. We defined dyslexia as $\leq 10^{\text{th}}$ %ile on a reading composite measure (i.e., the average of both subtests from the TOWRE A and TOWRE B) that measured the construct of single word and non-word reading accuracy and fluency. This cutoff identified 82 individuals as dyslexic and 727 as non-dyslexic.

Because the CLDRC sample is not a population sample of twins, we had to make adjustments to approximate a population sample. In the CLDRC, twin pairs are assigned to one of two groups based on school history. Group 1 (affected) includes twin pairs in which at least one member of the twin pair has a school history of ADHD, dyslexia, or math disability (MD). Group 2 (controls) includes only pairs where neither twin has a school history of ADHD, dyslexia, or MD. Although this sample is enriched for children with learning difficulties, school history provides a weak selection for group 1 (i.e., learning disabled twins have a lower overall mean than in population samples but a wide range of performance), and conversely, a weak selection for the absence of learning problems for group 2 (mean performance is typically higher in this group than population samples). To approximate a population sample in the CLDRC, we used all of the Group 2 (control) pairs, and 45% of the Group 1 (affected) pairs selected randomly. Earlier analyses have shown that these proportions result in an approximate population sample. So, in the CLDRC, we began

with 896 Group 1 pairs and 455 Group 2 pairs, and then randomly reduced Group 1 to 372 pairs, so that 45% of the combined sample came from Group 1 and 55% came from Group 2. We then randomly selected a single twin from each pair resulting in a final sample of 827 individuals. We defined dyslexia as $\leq 10^{\text{th}}$ %ile on a reading composite that mainly captured single word reading accuracy (PIAT Reading Recognition subtest and the Time Limited Word Recognition Test), resulting in a sample of 83 dyslexic individuals and 744 non-dyslexic individuals.

Demographic variables and scores on standard reading measures and IQ are given in Table 2, where it can be seen that each sample approximates a population sample. The mean age of each sample is different, with the ILTS having a lower mean as well as a narrower range than the CLDRC sample. In the ILTS sample, cognitive predictors were measured when the children were in Kindergarten, and reading outcome was measured when the children were in 1st grade. The mean age of the ILTS sample reflects their second time point (when reading was assessed). In the CLDRC sample, participants ranged in age from 8–16 years, and cognitive predictors of reading and reading outcome were measured concurrently.

Measures

The measures of the three cognitive constructs, PA, L, and PS or NS, used to predict reading skill in each sample are shown in Table 3. There were multiple measures of each construct and previous confirmatory factor analyses in each sample yielded the hypothesized three factors on which these individual measures loaded. It should be noted that there were only naming speed measures available in the ILTS sample, whereas the CLDRC PS/NS composite score included both naming speed and perceptual speed measures, like WISC-III Symbol Search and Coding. A previous study of these measures in the CLDRC sample (Shanahan et al., 2006) found deficits in dyslexia on both the naming speed and perceptual speed factors and that these factors were highly correlated ($r = .76$). So, in the current study, we combined naming and perceptual speed measures in the CLDRC sample to create a single composite score. The procedures followed to administer these measures to each twin sample are described in previous publications (Byrne *et al.*, 2002; DeFries et al., 1997).

Data Analysis Plan

Two methods were used to test the fit of single and multiple deficit models of dyslexia to the data. The first method is the familiar “counting deficits” method employed by Ramus et al. (2003), Nigg et al. (2005) and others, and the second method is regression fit. This first method used an arbitrary cutoff of 10th percentile (based on the control mean) to determine the presence of a deficit in PA, VC, and PS/NS. We then counted how many deficits individuals with and without dyslexia had. As explained earlier, the five theoretical models tested in this paper make different predictions about how many deficits those in the dyslexic (and non-dyslexic groups) will have, and whether the majority of dyslexic cases will have a deficit in PA. While this method of counting deficits readily identifies which is the most prevalent *deficit* profile in the group with the diagnosis, and how many in that group have single vs. multiple deficits, it does not provide a very stringent test of whether an individual fits a particular model. An individual with dyslexia may have three deficits and another may have no deficits, yet the reading score of each individual may well be explained by the same regression model (with one, two, or more cognitive predictors). This is true because additional deficits beyond a core deficit may be incidental in explaining the reading performance in a dyslexic individual with multiple deficits. Similarly, an individual with mild dyslexia may have no cognitive predictors below the deficit threshold (e.g. the 10 %ile of the population), but may still be a good fit to a given single or multiple predictor regression model.

To deal with the issue of whether an individual's reading skill is proportional to the cognitive deficits they have, we used a second, regression-based method to evaluate the fit of an individual to competing regression equations. We used four different regression equations in this second method: three single predictor equations (with either PA, PS/NS, or L predicting reading skill) and the best multiple predictor equation (with the optimal combination of PA, PS/NS, and L). We then tested which of the four regression equations best predicted an individual's reading score (i.e. yielded the lowest standardized residual for each individual). In this second method, every individual case necessarily "fit" one of the regression equations, as they were assigned to one of them based on which yielded the smallest standardized residual. We then combined the two methods, counting deficits and regression fit, to see which dyslexic and non-dyslexic cases were a good fit to a given model using both criteria. In this way, we could be fairly certain that the relative success of one model over another was robust across methods.

We proceeded to test the diagnostic utility of different single and multiple deficit profiles for identifying dyslexia in each sample using statistics used to evaluate screening measures: sensitivity (the proportion of actual cases that were correctly predicted, that is the ratio of true positives to the total of all cases), specificity (the proportion of non-cases that were correctly predicted, that is the ratio of true negatives to the total of non-cases), positive predictive power (PPP, the ratio of true positives to all positives), and negative predictive power (NPP, the ratio of true negatives to all negatives).

We were also interested in whether our cognitive predictors were just proxies for a demographic variable like parent education or for the well-known familiarity of dyslexia. If either were true, then it would be more efficient for clinicians and educators to use those variables to predict which children are at risk for dyslexia rather than test children on the various cognitive predictors. To address this question, we used hierarchical regressions to test the strength of the cognitive predictors after the contribution of either parent education or parent reading history was accounted for. Finally, secondary analyses were conducted to explore the effects of age, language, and comorbid ADHD status on the results from the primary analyses.

Results

The results of our analyses are organized into four main sections: 1) regression equations predicting reading skill, 2) prediction of individual cases using both methods (counting deficits and best fit to a regression equation), 3) clinical utility of different deficit profiles, and 4) secondary analyses looking at age, language, and ADHD effects.

Regression equations predicting reading skill

We fit single predictor regression equations to each cognitive predictor in each sample. We then determined the best multiple predictor equation for each sample. As can be seen in Table 4, the results for the single regression equations were similar in the two samples for PA (which was the strongest predictor in each) and for PS/NS. The results differed for language skill (L), which was a considerably stronger predictor of reading skill in CLDRC than in ILTS. Turning to multiple predictor regression equations, in the CLDRC sample, the best fitting multiple predictor equation for reading skill included all three cognitive predictors (PA, L, and PS/NS) and the interaction between PA and L (Table 5). As can be seen in Table 5, the strongest cognitive predictor was PA, followed by L, then PS/NS and then the PA \times L interaction. This equation accounted for 67% of the variance in reading skill in the CLDRC, which approached the maximum possible prediction of 74.6%, given the reliabilities of the measures in the equation. To determine the direction of the interaction, we created four subgroups by median splits on the two interacting predictors, PA and L, and

plotted the reading scores of each group. We did this first in the entire sample and then in the non-dyslexic group. This procedure revealed that reading skill in the non-dyslexic group was disproportionately good for the subgroup that was high on both predictors; thus, the interaction applied to the prediction of good reading, not poor reading.

As described earlier, we wanted to know if these cognitive predictors were just proxies for either parent education or parent history of reading problems. To answer this question, we ran four hierarchical regressions in the CLDRC sample, entering the individual parent variables either first in Block 1 or second in Block 2 after the cognitive predictors. Parent education accounted for 13.6% of the variance when entered first and less than 1% when entered second. So, most of the relation between these cognitive predictors and reading skill is independent of parent education. For parent reading history, the prediction was even weaker. Parent reading history accounted for 7.8% of the variance in reading skill when entered first and was not significant when entered second.

The best fitting multiple predictor model in the ILTS included only two of the three predictors, PA and NS (in that order), followed by the interaction of PA by NS. Language skill (L) was not a significant predictor. As can be seen in Table 5, this model accounted for 51.9% of the variance in reading skill, which is 22% less than the maximum possible prediction of 74% given the reliabilities of the measures in the equation. This lower value is to be expected given that the prediction was longitudinal from kindergarten to first grade (vs. concurrently measured as in the CLDRC sample), and given that the sample was younger than the CLDRC sample. Again, to understand which groups may be driving the interaction, we plotted mean reading scores of subgroups defined by median splits on PA and NS. Once again, the interaction occurred at the high end of reading skill, with the subgroup of non-dyslexic individuals who were high on both predictors having a disproportionately good reading score. So, across both samples, an interactive model of dyslexia was not supported; instead, the multiple predictors made *additive* contributions to predicting dyslexia.

As in the CLDRC sample, we next examined the relation between these cognitive predictors of reading skill and parent education in the ILTS sample (no parent history of reading difficulty was administered in the ILTS sample). Parent education was a weaker predictor of later reading skill in this sample, accounting for 2% of the variance when entered first and no significant variance when entered second.

In terms of single predictor regression equations, the results from the CLDRC sample are, as expected, consistent with the multiple regression results. The best single predictor model was that for PA (54.6% of the variance in reading skill), followed by L (40%), and then PS/NS (34.8%), with each equation accounting for appreciably less than the maximum set by the reliabilities, roughly 73–76% (see Table 4).

The results in the ILTS sample also showed consistency across single and multiple predictor regression equations, as expected. The best single predictor regression equation was that for PA (47.6% of the variance in reading skill), followed by NS (27.4%), and then L (10.9%), with each equation accounting for considerably less than the maximum set by the reliabilities (see Table 4).

In sum, the results of the regression equations are generally similar across samples. The main findings can be summarized as follows: PA is the best single predictor in each sample; a multiple predictor model accounts for more variance in reading skill than any single predictor model; synergy among predictors occurs for better than average reading rather than below average reading; and most of the cognitive prediction of reading skill is independent of parent education (and parent reading history in CLDRC). The primary

difference between the two samples is that the results for L are considerably weaker in the ILTS sample than in the CLDRC sample. This is likely due to the fact that participants were younger in the ILTS sample and the difference in language measures used across samples.

With regard to the five competing models of reading skill described above, the regression results clearly reject the first model, the single phonological deficit model, which has been the dominant model of dyslexia in the literature, given the fact that the multiple predictor regression equations predict a greater amount of reading variance overall. The remaining four models are all consistent with these regression results. To choose among them, we need to examine their success at predicting individual cases.

Prediction of Individual Cases

As described earlier, we used two different methods to evaluate the fit of individual cases to a given model: 1) presence of the predicted deficit and 2) best fit to that regression equation. We cross-tabulated the results of these two methods in each sample. We will briefly review the predictions the five theoretical models discussed earlier make for these cross-tabulated results. Model 1 (single phonological deficit) predicts that the large majority of cases will have a single phonological deficit and fit the single PA regression equation. Model 2 (single deficit subtypes) predicts that the large majority of cases will have a single deficit (in PA, L, or PS/NS) and fit the corresponding regression equation. Model 3 (phonological core, multiple deficit model) predicts the large majority of cases will have at least two deficits, one being a PA deficit, and will fit the multiple predictor regression equation (which includes PA as the strongest predictor in each sample). Model 4 (multiple deficit model, multiple predictor model) predicts the large majority of cases will have at least two deficits and fit the multiple predictor regression equation, but does not require a PA deficit. Model 5 (hybrid model) predicts some cases will fit Model 2 and others will fit Model 4.

CLDRC results—As can be seen by tallying the bolded entries in Table 6, 46% of the dyslexic cases (38 out of 83) satisfied both methods of individual prediction for a given model using a deficit cutoff set at the 10th %ile. Thus, only about half of the CLDRC dyslexic group was a good fit to a given model. Being a good fit meant that they had *only* the predicted deficit or deficits and that they best fit the corresponding regression equation, based on a comparison of their standardized residuals. Considering our five theoretical models in turn, we can quickly reject Model 1 (single phonological deficit), because only 13% (11/83) of cases fit that model. We can also reject Model 2 (single deficit subtypes) because only 24% (20/83) cases fit that model. Model 3 (phonological core, multiple deficit, multiple predictor model) can also be rejected because only 20% (17/83) cases fit that model. And Model 4 (multiple deficit, multiple predictor model) was only slightly better than Model 3, accounting for 22% (18/83) cases. So none of these models comes at all close to accounting for a majority of cases. That leaves the hybrid model, which encompasses models 2 and 4, and so accounts for all 38 cases that were counted as a good fit (38/83 or 46% of the entire sample).

Of the remaining 45 cases that did not satisfy both methods for model fit, the majority fit the single predictor PA regression model but had either multiple or no deficits. Additionally, smaller numbers of cases fit the single predictor L or PS models, either with multiple deficits or no deficits. Of the 18 dyslexic cases without cognitive deficits (“false negatives”), seventeen had borderline scores (11th to 25th percentile) on one or more cognitive predictors (PA, L, or PS/NS). So a dyslexic without at least a borderline cognitive deficit is a rarity (1.2% of the sample).

The results for the non-dyslexic cases in the CLDRC sample are given in Table 7. As expected, the large majority (86%) of non-dyslexic cases had no deficits based on the 10th

percentile cutoff, but a small subset of non-dyslexic cases (19 or 2.6%) were a good fit to a model of reading disability, having both the requisite deficit and fitting the corresponding regression model. On inspection, these “false positive” cases had lower reading scores than the average for the non-dyslexic group, but did not meet the 10th percentile cutoff for dyslexia. The values in parentheses in Table 10 show the results for the non-dyslexic group when borderline cases (i.e. those with a reading score >10th percentile but <26th percentile) were eliminated. As can be seen, the false positive rate drops considerably, with only two cases (0.4%) having the requisite deficit (i.e. in L) and fitting that regression equation. Also, after eliminating these borderline cases, only 8.3% (51/612) have a deficit in PA, L, PS/NS or some combination. As was true for dyslexic cases, there is considerable heterogeneity among non-dyslexic cases in terms of which regression models they fit.

ILTS results—Generally, similar results were found for dyslexic cases in the ILTS sample, as can be seen in Table 8. A somewhat smaller proportion of the total (32 out of 82, 39%) satisfied both methods of individual prediction for a given model. Interestingly, 72% (23 out of 32) of the cases that satisfied both methods for good fit ended up fitting one of the single deficit models, a result which favors Model 2 (single deficit subtypes model). The remaining 50 dyslexic cases that were not a good fit were mainly predicted by the single PA or PS regression models, but either had multiple or no deficits. Of the 26 dyslexic cases with no deficits (“false negatives”), 23 had borderline scores (11th to 25th percentile) on one or more cognitive predictors. So there were only three dyslexics without at least a borderline cognitive deficit (3.7%). The main difference between the CLDRC and ILTS results is that the L predictor was much weaker in the ILTS sample, likely because of the difference in measures utilized.

The results for the non-dyslexic cases in the ILTS sample are given in Table 9. As in the CLDRC sample, the majority of non-dyslexic cases (601/727, 83%) had no cognitive deficits. But similar to CLDRC, there was a small subset of false positive cases (34/727 or 4.8%) who were a good fit to a reading disability model but did not have dyslexia. Again, on inspection, these “false positive” cases had lower reading scores than the non-dyslexic group mean, but did not quite meet the 10th percentile cutoff for dyslexia. The values in parentheses show how the results change when these borderline cases were eliminated. The false positive rate drops to 2.2% (13/603). Also, now only 13.6% (82/603) of non-dyslexics have a cognitive deficit, compared to 17.3% (126/727) in the entire non-dyslexic sample. Substantial subgroups also fit each regression model, with the PA model performing best (204 cases). Thus similar to CLDRC results, across both the dyslexic and non-dyslexic subgroups, the results generally support the hybrid model.

Clinical Utility of Cognitive Deficit Profiles

The predictive value of different cognitive deficit profiles in each sample is contained in Table 10, both with and without the borderline cases in the non-dyslexic group. It can readily be seen that single deficit models - PA only (Model 1) or either PA only, L only, or PS/NS only (Model 2 - single deficit subtypes) perform worse than the remaining models, all of which allow for more than one deficit. So, using any of these single deficit profiles has limited clinical utility. The best performing profile, shown in the last column of Table 10, is the “Any Deficit” profile. The latter profile corresponds to the hybrid model (Model 5 in the list presented earlier) because this profile allows for either a single or multiple deficit profile. The Any Deficit profile has similar sensitivity and specificity rates across both samples. (It can be seen that both specificity and PPP increased when borderline cases were removed.) Sensitivity was 78.3% in the CLDRC and 68.3% in the ILTS, while specificity was 86.2% (91.7% without borderline cases) and 82.7% (86.4% without borderline cases),

respectively. Because of the base rates of cases (10%) and non-cases (90%), sensitivity and PPP were inevitably lower than specificity and NPP across all profiles.

Since many clinicians and educators use a PA deficit to confirm a dyslexia diagnosis, we examined the utility of that practice in each sample. In the CLDRC sample, 55% (46/83) dyslexic cases had a PA deficit, either alone or in combination with other cognitive deficits. In the ILTS, the corresponding rate was 43% (35/82). So using the presence of a PA deficit either to screen for dyslexia or to confirm a dyslexia diagnosis would miss about half the cases of dyslexia.

We can compare the success of the “Any Deficit” profile with dyslexia (defined as the bottom 10% of the population) predicting itself from the end of first grade to the end of second grade in the ILTS sample. Dyslexia predicting itself also illustrates the limits on categorical prediction discussed earlier. To do this, we divided the dyslexic cases at the end of first grade into those whose dyslexia persisted at the end of second grade (56 out of 82 or 68%) and those whose dyslexia did not persist (26 out of 82 or 32%). We also determined how many non-dyslexic cases at the end of first grade became dyslexic at the end of second grade (23 out of 727 or 3.2%) and how many did not (704 out of 727 or 96.8%). These numbers yield a sensitivity of 70.9%, a specificity of 96.4%, a PPP of 68.3%, a NPP of 96.8%. Thus, we can see that the “Any Deficit” profile has a similar sensitivity and NPP to dyslexia predicting itself, but lower specificity and PPP.

Secondary analyses to explore effects of age, country, and ADHD status

Since the ILTS sample is comprised of participants tested in three different countries (Australia, U.S., and Norway), we wanted to ensure that the main results of this study were not different for children whose testing was conducted in Norwegian. To test this, we split the ILTS sample into two subgroups, with Australian and U.S. participants in one, and Norwegian participants in the other. The best fitting multiple cognitive predictor regression equation was then run in each subgroup, and the results compared to the findings for the entire ILTS sample. Results in the subgroups were very similar to each other, and consistent with findings from the entire ILTS sample. Specifically, the PA and NS standardized coefficients were statistically significant in each of the regression analyses, with the size of each beta commensurate with what was found in the overall sample. The only difference noted was in the Norwegian sample, where the PA \times NS interaction was not quite significant, which is likely due to reduced sample size.

A second important analysis looked at the possibility that the contribution of cognitive predictors to the prediction of reading might change with age (i.e., an effect of development). This could only be explored in the CLDRC sample, as only it had a substantial age range. Using a median split on age, the dyslexic cases in the CLDRC sample were split into younger vs. older subgroups (mean ages = 9.4 and 13.5 respectively). The best fitting multiple predictor regression model in the CLDRC sample (with PA, L, PS/NS, and PA \times L included as IVs) was run in each of the age groups. Results were very similar to those presented in Table 5 for the entire sample, with all the cognitive predictors contributing significant amounts of unique variance in the younger and older age subgroups. The only differences noted in these subgroup analyses were that a) PS/NS became the second most important predictor after PA in the younger group (with L falling to third), and b) the PA \times L interaction was only a trend in the older subgroup. Overall, the predictors of reading were similar across age.

A second analysis was also conducted to investigate the potential effect of age on model fit. We looked to see whether there were any significant differences in age in the subgroups of children who met different models. An ANOVA was conducted and results showed that

there were no differences in age ($F = 1.1, p = .35$) when comparing children who met only the PA, L, or PS/NS models, or met the multiple deficit model.

The last set of analyses investigated the effect of comorbid ADHD status on the primary results. We split the CLDRC sample into two subgroups, one where participants had ADHD and the other not. The best fitting multiple predictor regression equation was then run in each subgroup, to see if ADHD status affected the prediction of reading skill. Results indicated very similar results across ADHD status subgroups with regard to overall variance explained in reading skill, significance of the standardized coefficients and order of variable importance in predicting reading. In the ILTS sample, because ADHD rates were lower (because of missing ADHD ratings from teachers), we added continuous ADHD scores to the best fitting multiple predictor regression equation. In this way, we could test for main and interaction effects of ADHD in predicting reading skill. Similar to the CLDRC results, no ADHD effects were found. Overall, there was no indication that comorbid ADHD had a significant moderating effect on the results of this study.

Discussion

This study had two overarching and related goals. One was theoretical: to test the fit of single vs. multiple deficit models of dyslexia at the level of individual cases. The second was to see how predictions derived from theoretical models inform clinical practice. Contrary to our own expectations, we found that the hybrid model, rather than the multiple deficit model, best fit the data in both samples. We also found that the “Any Deficit” profile, which corresponds to the hybrid model, outperformed other profiles in predicting both current and future cases of dyslexia. In what follows, we discuss both the theoretical and clinical implications of these results.

In retrospect, it is now clear to us that the results we obtained were virtually inevitable, given that cognitive predictors besides a PA deficit had incremental validity in predicting reading skill in our regression equations, and given the statistical constraints on prediction accuracy discussed earlier. The predictors were only moderately correlated with each other, no single predictor reached its maximum possible correlation with reading skill, and the multivariate distributions in both samples met tests of multivariate normality. Given these results, it was inevitable that there would be different regions of multivariate space populated by partially overlapping subgroups of both dyslexic and non-dyslexic cases, but without any definite boundaries between subgroups. All the models except the hybrid model made a deterministic, causal claim: that a certain pattern of deficits were either necessary, sufficient, or both, in causing dyslexia. Only the hybrid model was probabilistic in its assumptions about the relations between deficits and outcomes.

So, our results indicate that the relation between predictors and reading skill are probabilistic not deterministic. This result is important for clinical practice because it means that a clinician should not require a child with dyslexia to fit a particular deficit profile or even to have any cognitive deficits in the constructs considered here. We still think testing for these cognitive deficits is important in the diagnosis of dyslexia, because it provides converging evidence for the diagnosis, particularly since dyslexics without at least borderline cognitive deficits were quite rare in both samples (1.2% in CLDRC and 3.7% in ILTS). If a child meets the operational definition of dyslexia used here ($\leq 10^{\text{th}}$ % on a reading fluency measure) and does not have at least a borderline cognitive deficit in PA, PS/NS, or L, then more examination is warranted.

There are similarities between each of our five models and models that are in the literature. Model 1, the single phonological deficit model, is similar to the model proposed by Ramus

et al. (2003) and to the broader phonological hypothesis that is the most widely accepted view of dyslexia. Our results reject a strict, single deficit version of the phonological hypothesis, but we did find that PA deficits were common in both dyslexic samples and had the highest sensitivity of any single cognitive deficit. We did not examine reading comprehension as an outcome, but it is likely that a different pattern of predictors would predict reading comprehension.

Model 2 has some similarities to Wolff and Bower's (1999) double deficit hypothesis, as does Model 4. The double deficit hypothesis holds that a deficit besides a PA deficit, in naming (or processing) speed (PS/NS), can be sufficient to cause dyslexia, and that those with a double deficit in both PA and PS/NS have a more severe form of dyslexia. The double deficit hypothesis also implies there are three subtypes of dyslexia: one with a PA deficit only, one with a PS/NS deficit only, and one with a double deficit. We evaluated how many of the dyslexic cases in both samples fell into these three subtypes. When we examined the numbers of dyslexic cases in 2x2 contingency tables counting cases with only a PA deficit, only a PS/NS deficit, or only a double deficit, we found limited support for the predictions of the double deficit hypothesis. Although there were dyslexic cases with single deficits in either PA (14 in CLDRC and 13 in ILTS) or PS/NS (5 in CLDRC and 18 in ILTS), and cases with a double deficit (13 in CLDRC and 9 in ILTS), more than half of the dyslexia cases across the two samples (55/83 in CLDRC and 42/82 in ILTS) did not fall into one of three subtypes.

Model 3 is similar to Stanovich's (1988) phonological core, variable difference model, but Stanovich's model does not make as strong a causal claim as model 3. We should give either Stanovich (1988) or Bishop & Snowling (2004) credit for having models that come closest to fitting the results we obtained. Our results are also quite similar to those obtained by Morris et al. (1998). These researchers used cluster analysis in a large dyslexic sample to identify seven multiple deficit subtypes of dyslexia, all but one of which had a PA deficit. But in contrast to our Model 1, none of these subtypes consisted of a single PA deficit. They interpreted their results as being consistent with Stanovich's (1988) phonological core, variable difference model.

Clinical Implications

As noted above, the models tested here are similar to those proposed by dyslexia researchers and thus are likely used explicitly or implicitly by clinicians, educators, and policy makers, to inform their clinical decision making. It is well recognized that actuarial prediction is superior to clinical prediction (e.g. Dawes *et al.*, 1989) and that findings based on a group with a disorder do not generalize to every individual in the group. Nonetheless, clinicians, educators, and policy makers are required to make diagnostic decisions about which individuals need intervention or prevention and which do not. Unlike researchers, they do not have the luxury of only dealing with groups and distributions.

In this paper, we have examined a disorder at the individual level about which more is known at the cognitive level of analysis than perhaps any other behaviorally-defined disorder, either in children or adults. Yet one clear clinical implication is that we cannot use any one of these cognitive profiles to rule in or rule out dyslexia, because of the heterogeneity of cognitive profiles among individuals with dyslexia and the probabilistic relation between cognitive deficits and dyslexia. For instance, one practice that is becoming popular in some public schools is to rule out dyslexia if the child does not have a PA deficit. Clearly, such a practice is not supported by the current results, as discussed earlier.

These results also need to be considered in relation to other reading prediction studies using data from these samples. In the CLDRC sample, (Hulslander *et al.*, in press) examined the

longitudinal prediction of reading skill from a mean age of 10y to a mean age of 16y in a subset of the CLDRC participants studies here. Because the reading latent trait stability was so high (.98), other cognitive predictors like PA or NS, did not add to the prediction of later reading skill. So, at least at later ages, reading predicts itself very well. Because the current study of the CLDRC sample was cross-sectional, and cognitive predictors as well as reading measures were given concurrently, we could not examine the incremental validity of cognitive predictors beyond reading predicting itself. In the ILTS sample, Furnes and Samuelsson (2010) demonstrated that preschool prediction of later reading skill varies somewhat by language. In the more orthographically transparent Norwegian and Swedish languages, the utility of PA as a predictor of reading skill was limited to the end of first grade, unlike the results in the two English speaking samples (from the USA and Australia). NS, in contrast, was similarly predictive across languages for both end of first and second grade reading. Our secondary analyses looking at the effect of English vs. Norwegian did not find significant differences in terms of the significance or relative importance of the cognitive predictors, but we only conducted our analysis with data from Kindergarten and 1st grade, and not at older ages. Finally, preschool letter knowledge was a strong predictor in both English speaking samples, consistent with many previous studies. We did not include letter knowledge in the current study because we were focused on oral language predictors of reading skill and because the letter knowledge measure had restricted range (because of ceiling effects) at the end of kindergarten when the other oral language measures were given. However, an ideal preschool screening battery (at least in English) would include letter knowledge as a predictor.

One reaction to our results and those of Hulslander et al., (in press) would be to abandon the use of these cognitive predictors in diagnostic evaluations of school age children for dyslexia (e.g. Fletcher *et al.*, 2007) and to instead just reliably assess reading skill itself. If the evaluation of these cognitive risk factors does not add anything to the diagnostic decision or the treatment plan, then this argument makes sense. However, we think this is a premature conclusion. Even if these cognitive predictors are not necessary or sufficient for the diagnosis of dyslexia, their sensitivity to the diagnosis (Table 10) is substantial, and so a diagnostician will be more confident of the diagnosis of dyslexia if it is also supported by evidence from these cognitive risk factors. Clinical diagnoses of dyslexia and other disorders should be based on converging evidence from the child's family and developmental history, qualitative observations of the child's behavior in the evaluation, and test scores (Pennington, 2009). As we have demonstrated, even reading tests are not perfectly reliable and children may perform poorly on them for a variety of reasons, and yet not have other converging evidence of a dyslexia diagnosis. Finally, when it comes to early identification and prevention of dyslexia, we obviously cannot measure reading skill in preschool children. Measures of these cognitive risk factors are currently among the best predictors we have, along with letter name (and sound) knowledge (e.g. Pennington & Lefly, 2001; Scarborough, 1998). A screening test composed of letter knowledge, PA, and NS or PS would be fairly quick and would help identify those children in need of early preventive intervention.

References

- Bishop, DVM. Test for reception of grammar. 2nd ed.. Abingdon, UK: Medical Research Council; 1989.
- Bishop DVM, Snowling MJ. Developmental dyslexia and specific language impairment: Same or different? *Psychological Bulletin*. 2004; 130(6):858–886. [PubMed: 15535741]
- Byrne B, Delaland C, Fielding-Barnsley R, Quain P, Samuelsson S, Høien T, et al. Longitudinal twin study of early reading development in three countries: Preliminary results. *Annals of Dyslexia*. 2002; 52:47–73.

- Christensen K, Petersen I, Skytthe A, Herskind AM, McGue M, Bingley P. Comparison of academic performance of twins and singletons in adolescence: Follow-up study. *BMJ*. 2006; 333(7578):1095. [PubMed: 17012267]
- Dawes RM, Faust D, Meehl PE. Clinical versus actuarial judgment. *Science*. 1989; 243(4899):1668–1674. [PubMed: 2648573]
- DeFries JC, Filipek PA, Fulker DW, Olson RK, Pennington BF, Smith SD, et al. Colorado learning disabilities research center. *Learning Disabilities: A Multidisciplinary Journal*. 1997; 8:7–19.
- DeFries JC, Singer SM, Foch TT, Lewitter FI. Familial nature of reading disability. *British Journal of Psychiatry*. 1978; 132:361–367. [PubMed: 638389]
- Denckla MB, Rudel RG. Rapid "automatized" naming of pictured objects, colors, letters and numbers by normal children. *Cortex*. 1974; 10(2):186–202. [PubMed: 4844470]
- Dunn, LM.; Markwardt, FC. Peabody individual achievement test. Circle Pines, MN: American Guidance Service; 1970.
- Fisher, JP.; Glennister, JM. The hundred pictures naming test. Hawthorn, Australia: Australian Council for Educational Research; 1992.
- Fletcher, JM.; Lyon, GR.; Fuchs, LS.; Barnes, MA. Learning disabilities: From identification to intervention. New York: New York: Guilford Press; 2007.
- Furnes B, Samuelsson S. Predicting reading and spelling difficulties in transparent and opaque orthographies: A comparison between scandinavian and us/australian children. *Dyslexia*. 2010; 16(2):119–142. [PubMed: 20440743]
- Golden, CJ. Stroop color and word test: A manual for clinical and experimental uses. Chicago, IL: Skoelting; 1978.
- Gough PB, Tunmer WE. Decoding, reading, and reading disability. *RASE: Remedial & Special Education*. 1986; 7(1):6–10.
- Groopman, J. How doctors think. New York, NY: Houghton Mifflin Co; 2007.
- Hulstlander J, Olson RK, Willcutt EG, Wadsworth SJ. Longitudinal stability of reading-related skills and their prediction of reading development. *Scientific Studies of Reading*. 14(2) (in press).
- IDA. Definition of dyslexia. Baltimore, MD: International Dyslexia Association; 2002. (www.Interdys.Org)
- Keating PA, Manis F. The keating-manis phoneme deletion test.
- Lindamood, CH.; Lindamood, PC. Lindamood test of auditory conceptualization. Pearson; 1979.
- McCarthy, JJ.; Kirk, SA. The illinois test of psycholinguistic abilities. Urbana, IL: University of Illinois Press; 1961.
- McGrath LM, Pennington BF, Shanahan MA, Santerre-Lemmon LE, Barnard HD, Willcutt EG, et al. A multiple deficit model of reading disability and attention-deficit/hyperactivity disorder: Searching for shared cognitive deficits. *Journal of Child Psychology and Psychiatry*. 2011; 52(5): 547–557. [PubMed: 21126246]
- Morris RD, Shaywitz SE, Shankweiler DP, Katz L, Stuebing KK, Fletcher JM, et al. Subtypes of reading disability: Variability around a phonological core. *Journal of Educational Psychology*. 1998; 90:347–373.
- Nigg JT, Willcutt EG, Doyle AE, Sonuga-Barke EJ. Causal heterogeneity in attention-deficit/hyperactivity disorder: Do we need neuropsychologically impaired subtypes? *Biological Psychiatry*. 2005; 57(11):1224–1230. [PubMed: 15949992]
- Pennington BF. From single to multiple deficit models of developmental disorders. *Cognition*. 2006; 101(2):385–413. [PubMed: 16844106]
- Pennington, BF. Diagnosing learning disorders. 2nd ed.. New York, NY: Guilford Press; 2009.
- Pennington BF, Bishop DV. Relations among speech, language, and reading disorders. *Annual Review of Psychology*. 2009; 60:283–306.
- Pennington BF, Lefly DL. Early reading development in children at family risk for dyslexia. *Child Development*. 2001; 72(3):816–833. [PubMed: 11405584]
- Peterson RL, Pennington BF. Seminar: Developmental dyslexia. (in review).

- Peterson RL, Pennington BF, Shriberg LD, Boada R. What influences literacy outcome in children with speech sound disorder? *Journal of Speech, Language and Hearing Research*. 2009; 52(5): 1175–1188.
- Ramus F, Rosen S, Dakin SC, Day BL, Castellote JM, White S, et al. Theories of developmental dyslexia: Insights from a multiple case study of dyslexic adults. *Brain: A Journal of Neurology*. 2003; 126(4):841–865. [PubMed: 12615643]
- Reitan RM. Trail making test results for normal and brain-damaged children. *Perceptual and Motor Skills*. 1971; 33:575–581. [PubMed: 5124116]
- Scarborough, HS. Early identification of children at risk for reading disabilities: Phonological awareness and some other promising predictors. In: Shapiro, BK.; Accardo, PJ.; Capute, AJ., editors. *Specific reading disability: A view of the spectrum*. Timonium, MD: York Press; 1998. p. 75-119.
- Stanovich KE. Explaining the differences between the dyslexic and the garden-variety poor reader: The phonological-core variable-difference model. *Journal of Learning Disabilities*. 1988; 21(10): 590–604. [PubMed: 2465364]
- Torgesen, JK.; Wagner, RK.; Rashotte, CA. *A test of word reading efficiency (towre)*. Austin, TX: Pro-ED; 1999.
- Wagner, RK.; Torgesen, JK.; Rashotte, CA. *Ctopp: Comprehensive test of phonological processing*. Austin, TX: PRO-ED; 1999.
- Wechsler, D. *Wechsler intelligence scale for children*. third edition. San Antonio, TX: The Psychological Corporation; 1991.
- Wechsler, D. *Wechsler preschool and primary scale of intelligence*. 3rd ed.. San Antonio, TX: Pearson; 2002.
- Wolf M, Bowers PG. The double-deficit hypothesis for the developmental dyslexias. *Journal of Educational Psychology*. 1999; 91(3):415–438.

Table 1

Four Cognitive Models of Dyslexia

		Single Deficit?	
		Yes	No
PA Deficit Necessary?	Yes	I: Single phonological deficit	III: Phonological core, multiple deficit
	No	II: Single deficit subtypes	IV: Multiple deficit

Table 2

Demographics for CLDRC (N=827) and ILTS (N=809)

Variable	Dyslexic		Non-dyslexic		Overall Sample	
	CLDRC (n=83)	ILTS (n=82)	CLDRC (n=744)	ILTS (n=727)	CLDRC (n=827)	ILTS (n=809)
Age	11.66 (2.31) [8.15 – 18.24]	6.2 (0.35) [5.3 – 6.8]	11.39 (2.57) [7.98 – 19.19]	6.3 (0.37) [5.3 – 7.7]	11.42 (2.55) [7.98 – 19.19]	6.3 (0.38) [5.3 – 7.7]
Parental Education ^a	13.73 (2.36) [9.0 – 20.0]	13.30 (2.18) [9.0 – 21.0]	15.11 (2.23) [9.0 – 21.5]	14.11 (2.30) [9.0 – 21.0]	14.98 (2.28) [9.0 – 21.5]	14.02 (2.30) [9.0 – 21.0]
Nonverbal IQ	96.02 (13.48) [46 – 128]	92.71 (13.95) [69.6 – 126.2]	106.84 (12.16) [68 – 147]	100.81 (14.90) [100.8 – 145.4]	105.75 (12.71) [46 – 147]	100.00 (15.00) [60.9 – 145.4]
Full Scale IQ	92.54 (12.39) [51 – 118]	-	109.34 (12.29) [67 – 142]	-	107.66 (13.29) [51 – 142]	-
Reading composite ^b	-1.91 (.42) [-3.0 to -1.44]	-	0.21 (.80) [-1.42 – 2.63]	-	0.00 (1.0) [-3.0 – 2.63]	-
TOWRE (z-score) ^c						
Australia	-	-1.43 (0.20) [-1.87 to -1.20]	-	0.18 (0.93) [-1.17 – 2.65]	-	0.04 (1.0) [-1.87 – 2.65]
United States	-	-1.62 (0.26) [-2.12 to -1.19]	-	0.17 (0.89) [-1.18 – 3.00]	-	-0.02 (1.0) [-2.12 – 3.0]
Scandinavia	-	-1.44 (0.23) [-1.86 to -1.19]	-	0.14 (0.96) [-1.15 – 2.76]	-	0.0 (1.0) [-1.86 – 2.76]
Sex	57.8% male	62.2% male	48.7% male	49.5% male	49.6% male	50.8% male

Note. Standard deviations are presented in parentheses. Ranges are presented in brackets. CLDRC = Colorado Learning Disabilities Research Center;

^aMean of mother and father years of education.

^bStandardized composite of PIAT Reading Recognition and Time-Limited Word Recognition.

^cTOWRE z-scores are presented for each country and were computed by regressing each raw score for the four TOWRE subtests (Sight Word Efficiency A & B and Phonemic Decoding Efficiency A & B), onto age and age-squared, saving the standardized residual. We then standardized these residuals *within* country to control for country differences. Finally, we averaged the resulting four z-scores to obtain the TOWRE z-score.

Table 3

Individual Measures of Each Construct

Variable	Samples	
	CLDRC	ILTS
PA	Phoneme Deletion 1; Phoneme Deletion 2; Pig Latin; and Lindamood	CTOPP: Elision, Blending Words, and Sound Matching
L	WISC-III: Information, Similarities, Vocabulary and Comprehension	Grammatic Closure; Morphology; WPPSI Vocabulary; 100 Pictures Naming Test; and Test for Reception of Grammar
PS/NS ^a	WISC-III: Symbol Search and Coding; Colorado Perceptual (CPS) Identical Pictures; and Trails A & B Rapid Automatized Naming (color, pictures, numbers and letters); Stroop baseline color and word naming	CTOPP: Rapid Naming Colors and Rapid Naming Digits
Reading	PIAT Reading Recognition; and Time Limited Word Recognition	TOWRE: Phonemic Decoding and Sight Word Efficiency (Forms A and B)

Note. CLDRC = Colorado Learning Disabilities Research Center; ILTS = International longitudinal Twin Study; PA = phoneme awareness; L = language skill; PS = processing speed; NS = naming speed.

^aComposite of PS & NS for CLDRC; composite of NS for ILTS.

Table 4
 Single Regression Equations for Each Sample, with Reading Skill as the Dependent Variable

Variable	Samples									
	CLDRC				ILTS					
	PA	L	PS/NS	PA	L	NS	PA	L	NS	
Theoretical Maximum R	.87	.86	.85	.85	.85	.69	.88			
R	.74	.63	.59	.69	.33	.52				
R ²	.55	.40	.35	.47	.11	.27				

Note. CLDRC = Colorado Learning Disabilities Research Center; ILTS = International longitudinal Twin Study; PA = phoneme awareness; L = language skill; PS = processing speed; NS = naming speed. Theoretical maximum R is based on the reliabilities of each of the predictors.

Table 5

Best-Fitting Multiple Regression Equations

Variable	β		SE		Δ R ²		P-value
	CLDRC	ILTS	CLDRC	ILTS	CLDRC	ILTS	
PA	.513	.540	.026	.029	.547	.292	<.001
L	.266	-	.025	-	.079	-	<.001
PS/NS ^a	.236	.280	.024	.031	.038	.078	<.001
PAXL	.073	-	.018	-	.007	-	<.001
PAXPS/NS	-	.074	-	.026	-	.005	<.010

Note. For CLDRC: Multiple R = .819, Adjusted R² = .670; For ILTS: Multiple R = .722, Adjusted R² = .569; CLDRC = Colorado Learning Disabilities Research Center; ILTS = International Longitudinal Twin Study; PA = phoneme awareness; L = language skill; PS/NS = composite of processing speed and naming speed.

^aComposite of PS & NS for CLDRC; composite of NS for ILTS

Table 6

Dyslexic cases in CLDRC: Cross tabulation of overall model fit, based on the “counting deficits” and the individual regression fit methods applies to individual cases

Deficit	PA only	L only	PS/NS only	PA and L	PA and PS/NS	L and PS/NS	PA, L, & PS/NS	None	Total
PA	11	0	0	1	7	0	3	8	30
L	0	5	0	1	0	1	2	1	10
PS/NS	0	0	4	0	0	4	1	4	13
Multiple	3	3	1	2	6	1	9	5	30
Total	14	8	5	4	13	6	15	18	83

Note. CLDRC = Colorado Learning Disabilities Research Center; PA = phoneme awareness; L = language skill; PS = processing speed; NS = naming speed.

Table 7

Non-dyslexic cases in CLDRC: Cross tabulation of overall model fit, based on the “counting deficits” and individual regression fit methods applied to individual cases

Deficit	PA only	L only	PS/NS only	PA and L	PA and PS/NS	L and PS/NS	PA, L, & PS/NS	None	Total
PA	2 (0)	6 (5)	8 (7)	0 (0)	1 (0)	2 (2)	0 (0)	167 (150)	186 (164)
L	6 (4)	7 (2)	4 (4)	3 (1)	2 (2)	1 (1)	3 (0)	160 (136)	186 (150)
PS	3 (2)	7 (6)	7 (0)	1 (1)	1 (0)	3 (1)	3 (1)	138 (112)	163 (123)
Multiple	10 (0)	12 (9)	7 (3)	1 (0)	1 (0)	1 (0)	0 (0)	177 (163)	209 (175)
Total	21 (6)	32 (22)	26 (14)	5 (2)	5 (2)	7 (4)	6 (1)	642 (561)	744 (612)

Note. CLDRC = Colorado Learning Disabilities Research Center; PA = phoneme awareness; L = language skill; PS = processing speed; NS = naming speed.

Table 8

Dyslexic cases in ILTS: Cross tabulation of overall model fit, based on the “counting deficits” and individual regression fit methods applied to individual cases

Deficit	PA only	L only	NS only	PA and L	PA and NS	L and NS	PA, L, & NS	None	Total
PA	10	0	0	5	2	0	0	12	29
L	0	2	0	0	0	0	0	1	3
NS	0	0	11	0	2	0	5	3	21
Multiple	3	0	7	1	5	1	2	10	29
Total	13	2	18	6	9	1	7	26	82

Note. ILTS = International longitudinal Twin Study; PA = phoneme awareness; L = language skill; PS = processing speed; NS = naming speed.

Table 9

Non-dyslexic cases in ILTS: Cross tabulation of overall model fit, based on the “counting deficits” and individual regression fit methods applied to individual cases

Deficit	PA only	L only	NS only	PA and L	PA and NS	L and NS	PA, L, & NS	None	Total
PA	10 (0)	8 (5)	7 (7)	3 (0)	4 (2)	1 (1)	2 (0)	169 (136)	204 (151)
L	6 (6)	15 (10)	6 (6)	0 (0)	2 (2)	2 (2)	3 (2)	118 (110)	152 (138)
PS	1 (1)	15 (15)	4 (2)	4 (4)	0 (0)	2 (0)	0 (0)	152 (133)	178 (155)
Multiple	7 (5)	8 (6)	11 (5)	2 (1)	2 (0)	0 (0)	1 (0)	162 (142)	193 (159)
Total	24 (12)	46 (36)	28 (20)	9 (5)	8 (4)	5 (3)	6 (2)	601 (521)	727 (603)

Note. ILTS = International longitudinal Twin Study; PA = phoneme awareness; L = language skill; PS = processing speed; NS = naming speed.

Table 10

Predictive Value (% or ratio) of Different Deficit Profiles

Variable	PA only	PA	L only	L	PS/NS only	PS/NS	Any Single	2-3 Deficits	Any Deficit
CLDRC									
Sens	16.9	67.5	9.6	39.8	6.0	47.0	32.5	45.8	78.3
Spec	97.2	95.0	95.7	93.3	96.5	94.1	89.4	96.9	86.3
PPP	40.0	60.2	20.0	39.8	16.1	47.0	25.5	62.3	38.9
NPP	70.0	83.6	26.7	53.2	26.3	65.0	39.1	80.8	56.0
ILTS	91.3	96.3	90.5	93.3	90.2	94.1	92.2	94.1	97.3
ILTS	89.8	95.7	88.7	92.1	88.4	93.1	91.1	93.1	96.9
Sens	15.9	41.5	2.4	19.5	22.0	42.7	40.2	28.0	68.3
Spec	96.7	93.5	93.7	90.9	96.1	93.5	86.5	96.0	82.7
PPP	35.1	42.7	4.2	19.5	39.1	42.7	25.2	44.2	30.8
NPP	52.0	59.6	5.3	25.8	47.4	54.7	32.7	62.1	40.6
ILTS	91.0	93.5	89.5	90.9	91.6	93.5	92.8	92.2	95.8
ILTS	89.5	92.4	87.6	89.4	90.1	92.4	91.6	90.9	95.2

Note. CLDRC = Colorado Learning Disabilities Research Center; ILTS = International Longitudinal Twin Study; Sens = Sensitivity; Spec = Specificity; PPP = Positive Predictive Power; NPP = Negative Predictive Power – all expressed as percentages. All subject are included in the first row of each variable, and borderline subjects have been removed from the second row.