DYSLEXIA
Published online 13 June 2006 in Wiley InterScience
(www.interscience.wiley.com) DOI: 10.1002/dys.324

Meta-analysis of the Nonword Reading Deficit in Specific Reading Disorder

Julia A. Herrmann[1,*], Tom Matyas[1] and Chris Pratt[2]

[1] *School of Psychological Science, La Trobe University, Australia*
[2] *Professorial Unit, The Melbourne Clinic, Australia*

A meta-analysis was conducted to investigate whether specific reading disorder (SRD) groups demonstrate a deficit in using phonological recoding strategies. Thirty-four studies were reviewed that had compared the nonword reading performances of SRD groups with reading-level matched (RL) control groups. The average nonword reading difference between groups across the total number of studies was moderate ($d = 0.65$, $N = 2865$). Three predictors of the size of group differences in nonword reading ability were identified. Studies that used passage reading tests to match groups for reading level found significantly less evidence for nonword reading deficits than studies that used word-level reading accuracy tests. Secondly, there was a significant positive relationship between group differences in intelligence level (SRD-RL control group) and effect sizes. Finally, group differences in age showed a significant negative association with effect magnitudes. The mean age, reading level and intelligence level of groups did not significantly predict nonword reading outcomes. It was concluded that there was evidence for nonword reading deficits in SRD groups, consistent with the claim that deficient development of phonological recoding strategies is a leading cause of reading difficulties. Copyright © 2006 John Wiley & Sons, Ltd.

Keywords: specific reading disorder; meta-analysis; phonological recoding deficit; reading level matched control group; outcome predictors

INTRODUCTION

Specific reading disorder (SRD) is defined as a reading difficulty that occurs in the context of adequate intelligence and educational opportunity, and an absence of known sensory, neurological or behavioural problems that

*Correspondence to: Julia A. Herrmann, Older Persons and Rehabilitation Service, Waikato Health, PO Box 934, Hamilton, New Zealand. Tel.: 64-7-8398899; juliaherrmann@hotmail.com

would otherwise account for the difficulty (DSM-IV, American Psychiatric Association, 1994). In order to understand why this difficulty occurs, researchers have investigated whether the reading development of SRD groups has been arrested due to deficits in particular skills. The *deficit hypothesis* predicts qualitative differences between the reading development of children with SRD and normally developing readers. This prediction contrasts with the *developmental delay hypothesis*, which postulates that children with SRD acquire the same reading skills as normally developing readers, but at a slower rate.

Most individuals with SRD demonstrate a primary difficulty with printed word identification and reading accuracy at the individual word-level. SRD profiles characterized by adequate word-level reading skills but poor reading comprehension appear to be relatively rare (Shankweiler *et al.*, 1999; Snowling, 2002; Vellutino, Fletcher, Snowling, & Scanlon, 2004). Research evidence indicates that difficulties with printed word identification in SRD groups can be linked to problems using phonological recoding strategies. *Phonological recoding* involves 'sounding out' printed words using knowledge of letter–sound relationships. Nonword reading accuracy provides an index of the success with which unfamiliar words can be read aloud using phonological recoding strategies. Nonwords are meaningless, novel letter strings that have conventional spelling patterns, such as *tob*, *yone* and *migwup*.

The importance of skill in phonological recoding, measurable by nonword reading accuracy, is evident in most theoretical models of reading. There is a common assumption that a mechanism involved in mapping letter–sound relationships is necessary to account for reading behaviour in an alphabetic orthography such as English. For example, the dual-route model of skilled reading posits a nonlexical procedure, in which a string of letters is converted into a set of sounds using knowledge of regularities of relationships between letters and sounds (e.g. Coltheart, Rastle, Ziegler, & Langdon, 2001). In another leading model of skilled reading, the parallel distributed processing (PDP) connectionist model, the reading system is described as a knowledge base of the frequency and consistency of letter–sound patterns gained from exposure to printed words (e.g. Plaut, McClelland, Seidenberg, & Patterson, 1996).

The importance of phonological recoding strategies is also emphasized in developmental models of reading. Stage-based models posit that reading development is characterized by an increase in knowledge of letter–sound connections. Most strategies used by the developing reader involve phonological recoding; it is only at the last stage of reading development that children can identify words accurately using a different strategy, which is based on recognizing words as whole orthographic units (e.g. Ehri, 1998; Frith, 1985). In the self-teaching model, phonological recoding is conceptualized as a systematic and reliable method of generating pronunciations for the majority of words in the English language, and provides the opportunity to acquire representations of word-specific orthographic units (Share, 1995).

Given its centrality to reading, and reading development in particular, impairment in phonological recoding could offer a powerful explanation of SRD. A popular technique for evaluating the evidence for phonological recoding deficits is the reading-level matched group design, in which SRD groups are compared to groups of younger, normally developing readers of the same reading level. Groups are matched usually for mean performances on

DOI: 10.1002/dys

standardized reading accuracy tests. If SRD groups perform significantly below control groups of equivalent reading ability on a particular type of reading task, such as nonword reading, then the strategies underlying that task could be a potential skill deficit.

Reading-level matched design studies have provided an abundance of evidence for phonological recoding deficits in SRD, and at least two reviews of this literature have been published. The narrative review by Rack, Snowling, and Olson (1992) examined 16 studies published between 1980 and 1989 that had investigated whether SRD and control groups differ in phonological recoding skill despite equivalent reading abilities. The majority of these studies found that SRD groups attained significantly lower nonword reading performances than reading-level matched control groups, consistent with the deficit hypothesis. In these studies, the median group difference in nonword reading was 19 percentage points, suggesting a moderate deficit. Six of the studies reviewed, however, found that groups were equivalent in nonword reading consistent with the developmental delay hypothesis. Methodological issues were explored to reconcile the differences in findings across studies. Based on the overall evidence, however, Rack *et al*. (1992) concluded that most SRD groups demonstrate a phonological recoding problem beyond expectations for reading level, and that deficient phonological recoding skills are a likely source of reading difficulties.

Ijzendoorn and Bus (1994) subsequently published a meta-analysis of the 16 papers examined in Rack *et al*.'s (1992) narrative review. Using data from 1183 participants, the overall combined effect size for group differences in nonword reading ability was $d = 0.48$. That is, on average, the nonword reading performances of SRD groups fell half a standard deviation below expectations for reading level, a highly significant result.

Ijzendoorn and Bus (1994) reported strong variations in effect sizes among the studies reviewed, ranging from $d = 0.00$ to 1.03. Potential moderator variables were derived from the methodological issues raised by Rack *et al*. (1992), and several significant predictors of outcome were found. It was reported that between group differences in age, reading and intelligence levels correlated negatively with effect magnitude. Larger age differences between the groups were associated with smaller effect sizes. The reading-level matched group design involves comparison between two groups with different chronological ages. Ijzendoorn and Bus found that the greater the discrepancy in group ages, the lower the likelihood of detecting nonword reading group differences. Smaller group differences in nonword reading were also associated with higher levels of reading and measured intelligence in SRD groups compared to control groups. Furthermore, the types of tests used in reading-level matched design studies influenced findings. Matching groups for verbal intelligence test scores yielded larger group differences in nonword reading accuracy than matching for full-scale or nonverbal intelligence scores. Finally, those studies that used passage reading measures or the Woodcock tests to match groups for reading level had smaller effect sizes than studies that used other reading tests.

Ijzendoorn and Bus (1994) concluded that findings of large group differences in nonword reading were associated with the following study design procedures: use of verbal intelligence measures, use of reading tests other than passage reading tests or the Woodcock tests, and a close group match for age, reading and intelligence levels. Several methodological issues raised by Rack *et al*. (1992) had

DOI: 10.1002/dys

no discernable effect on study outcomes. These included the mean chronological ages of groups; use of simple versus complex nonwords as the dependent variable; and whether or not nonwords were visually similar to words. There was also no association between method of participant recruitment (e.g. learning disorder clinics versus regular primary schools) and study outcome.

These reviews provide valuable information about aggregated study outcomes and clear support for the deficit hypothesis. However, some criticisms of the proposed moderator variables can be noted. First, in Ijzendoorn and Bus' (1994) analysis, studies that used the Woodcock Word Identification tests or passage reading tests were grouped together because these tests were considered less adequate, and so formed one category. The Woodcock and passage reading tests have been criticized for different reasons, however, and it is unclear whether use of one or both tests lead to findings of smaller group differences in nonword reading ability. Rack *et al.* (1992) recommended against the use of passage reading measures because some children with SRD might use the context of the story passage to guess the identity of a word that they do not recognize. In contrast, Rack *et al.* criticized use of the Woodcock Word Identification tests, which measure individual-word level reading accuracy, because most items are regular words that can be recoded phonologically. Groups matched on predominantly regular word reading tests might have similar levels of phonological recoding skill.

Secondly, the direction of the effect of group differences in intelligence scores reported by Ijzendoorn and Bus (1994) contradicted the data trends reported by Rack *et al*. (1992). Ijzendoorn and Bus reported that the higher the intelligence of the SRD groups relative to controls, the smaller the effect size. They suggested that SRD groups with higher levels of intellectual functioning might have better nonword reading outcomes. In contrast, Rack *et al.* observed that those studies that did *not* find nonword reading deficits tended to have SRD samples with *lower* measured intelligence than control samples. Furthermore, Ijzendoorn and Bus examined the effect of group differences in intelligence scores only. It would be of particular interest to examine whether nonword reading outcomes across studies are influenced by differences in the measured intelligence levels of SRD samples.

Rack *et al.'s* (1992) recommendation to sample only SRD participants who show adequate verbal intelligence also requires further consideration. It was proposed that individuals with age-appropriate verbal intelligence were more likely to be genuine cases because their reading disorder could not be explained by general language problems. Rack *et al.* suggested that phonological recoding difficulties might be a particular feature of these 'genuine, unexplained' cases, and meta-analytic findings supported this idea.

In conducting research with the reading-level matched design, however, researchers have warned against using too many exclusionary criteria (Jackson & Butterfield, 1989). Research outcomes from samples defined by numerous exclusionary criteria may neither replicate easily, nor be representative of the broader SRD population. Furthermore, relatively low scores on verbal intelligence tests have been found in some SRD groups (D'Angiulli & Siegel, 2003; Snowling, 1991), raising the possibility that many children will be excluded from research studies if adequate verbal intelligence is a selection criterion.

In addition, the number of reading-level matched design studies has doubled since the review papers by Rack *et al*. (1992) and Ijzendoorn and Bus (1994) were

DOI: 10.1002/dys

published over 12 years ago. The increase in published work attests to the continued research interest in this aspect of reading behaviour. The majority of these studies have found that SRD groups performed significantly below reading-level control groups in nonword reading performances. Once again, however, a small number of recent studies found no evidence for nonword reading deficits in SRD groups (Pennington, Van Orden, Smith, Green, & Haith, 1990; Snowling, Goulandris, & Defty, 1996; Stothard & Hulme, 1995; Vellutino, Scanlon, & Tanzman, 1994).

Evidence for the nonword reading deficit has been obtained across sampling and testing variations, such as differences in chronological age and use of different nonword stimuli, consistent with the outcomes from Ijzendoorn and Bus' (1994) meta-analysis. Most recent studies have examined child and adolescent samples of the SRD population, which was the age range investigated in previous reviews. There is now evidence, however, for nonword reading deficits in adults with SRD as well (e.g. Greenberg, Ehri, & Perin, 1997).

Some recent findings appear to contradict arguments about predictors of the nonword reading deficit. Many studies that used the Word Identification subtest of the Woodcock tests reported that SRD groups performed significantly below reading-level control groups in nonword reading performances (e.g. Badian, 1997; Bowey & Hansen, 1994; Manis, Seidenberg, Doi, McBride-Chang, & Peterson, 1996). Furthermore, nonword reading deficits have been found in studies that used verbal intelligence measures (e.g. Bowey, Cain, & Ryan, 1992; Felton & Wood, 1992; Greenberg *et al*., 1997) and full-scale intelligence tests (e.g. Conners & Olson, 1990; Duncan & Johnston, 1999; Manis, Custodio, & Szeszulski, 1993; Murphy & Pollatsek, 1994). Hence, use of the Woodcock Word Identification tests to measure reading level does not preclude finding evidence for the nonword reading deficit in SRD samples, and there may be no particular association between the type of intelligence measure used and nonword reading outcomes, as previous arguments would suggest.

In summary, given the concerns outlined regarding Ijzendoorn and Bus' (1994) research on predictor variables of nonword reading outcomes, and taking account of the substantial number of research studies published subsequent to this, a further meta-analysis on the nonword reading deficit in SRD groups was warranted. The first aim of the meta-analysis presented in this study was to obtain an updated estimate of the effect size index for the difference in nonword reading ability between SRD and reading-level control groups, taking advantage of additional published work. Consistent with Ijzendoorn and Bus' findings, strong variations in effect sizes were expected. Therefore, a second aim was to examine heterogeneity of study outcomes, to gain information about the variables that influence study outcomes, and to determine whether the addition of new data would alter Ijzendoorn and Bus' conclusions.

METHOD

Selection of Studies

The set of 16 studies reviewed by Ijzendoorn and Bus (1994) previously was included in the meta-analysis. To qualify for analysis, any additional studies had

DOI: 10.1002/dys

to meet the following criteria:

(a) Nonword reading was used as a measure of phonological recoding skill.
(b) The studies used the reading-level matched group design.

In conducting an electronic search of the PsychInfo database, terms that referred to nonword reading were crossed with terms related to the relevant groups. The nonword reading terms used were: nonword reading, nonwords, pseudowords, phonological recoding, phonological reading, and phonological decoding. The terms entered for the SRD group were: SRD, reading disabilities, dyslexia, reading delays, reading difficulties, poor reading. The terms for the reading-level control group were: reading level control group, RL control group, reading age control group, RA control group, reading level, and reading grade.

Eighteen studies met the inclusion criteria. When added to the 16 studies examined previously, 34 studies were included in the meta-analysis. The studies reviewed by Ijzendoorn and Bus (1994) will be referred to as the *original studies*, whereas the more recent studies will be referred to as the *additional studies*. Several studies reported outcomes for multiple independent groups. There were 18 between-group comparisons in the original studies, and 21 in the additional studies. Each comparison represented an independent test of the nonword reading deficit hypothesis. These comparisons involved 2865 participants.

In several studies, more than one measure of the dependent variable (nonword reading ability) was administered for between-group comparisons. For the original studies, the same dependent variable measure was used that had been selected for analysis by Ijzendoorn and Bus (1994). For the additional studies, however, the average group performance on multiple nonword reading tasks was used because this outcome would be more reliable statistically than an outcome from a single reading measure.

The studies included in the meta-analysis are presented in Table 1. Data for sample size, age in years, reading level, mean intelligence standard score, and mean nonword reading performances are presented. Standard deviation data are presented in parentheses. Instances where studies reported reading level as a grade equivalent (Gr) or percentile rank (PR) are indicated; otherwise, data are reported as reading age in years. Nonword reading scores are percent correct unless indicated as a grade equivalent (Gr) or other type of measurement unit. It is noted that one older study (Frith & Snowling, 1983) is listed under the additional studies in Table 1. This study was not examined by Ijzendoorn and Bus (1994), and it met the inclusion criteria for this meta-analysis. All other additional studies were published between 1990 and 1999.

Descriptive data from the studies in Table 1 indicates that most SRD samples were reading at a level at least 2 years below their chronological age or grade placement. The control groups, in contrast, had reading levels commensurate with age or grade expectations. In most studies, the measured intelligence levels of groups fell within the average range, with a standard score of at least 80 or 90.

Data Retrieval and Meta-analytic Procedures

One of the aims of the meta-analysis was to determine whether findings from additional studies would alter the original conclusions made by Ijzendoorn and

DOI: 10.1002/dys

Table 1. Summary of studies comparing SRD and reading-level matched control groups on measures of nonword reading

Study	n	Age	Reading level	IQ	Nonwords
1. Original studies					
Snowling (*1980*)					
SRD	18	12.1	7–10	106.0	0.63*
Control	36	9.5	7–10	105.0	1.85
Snowling (*1981*)					
SRD	20	9.5–17.3	8–11	108.0	75.0
Control	22	7.2–10.2	8–10	94.0	88.5
Baddeley, Ellis, Miles, and Lewis (*1982*)					
SRD	15	12.8	10.3	108.0	58.4
Control	15	9.9	10.3	113.0	67.6
DiBenedetto, Richardson, and Kochnower (*1983*)					
SRD	20	10.2 (1.2)	Gr 3.6 (1.0)	111.3 (15.6)	70.0 (14.0)
Control	20	8.0 (0.7)	Gr 3.7 (1.1)	112.6 (14.7)	83.0 (16.4)
Kochnower, Richardson, and DiBenedetto (*1983*)					
SRD	20	10.3 (1.4)	Gr 3.6 (1.1)	108.0 (13.4)	48.6 (19.5)
Control	20	8.0 (0.8)	Gr 3.6 (1.0)	110.0 (11.7)	62.2 (19.3)
Beech and Harding (*1984*)					
SRD	57	9.9 (1.4)	7.6 (1.0)	97.8 (10.2)	29.6 (26.3)
Control	44	7.2 (1.1)	7.7 (1.3)	100.4 (8.6)	38.5 (32.2)
Olson, Kliegel, Davidson, and Foltz (*1985*)					
SRD	50	15.3	Gr 6.7	>90	70.0
Control	47	10.1	Gr 7.0	>90	81.0
Treiman and Hirsh-Pasek (*1985*)					
SRD	37	11.8 (1.5)	Gr 3.57 (1.2)	>80	64.2 (26.2)
Control	37	8.5 (1.1)	—	—	64.5 (25.3)
Johnston, Rugg, and Scott (*1987*)					
Younger[a1]					
SRD	20	8.5 (0.3)	7.2 (0.4)	96.8 (6.3)	35.0 (21.9)
Control	20	7.2 (0.2)	7.5 (0.4)	101.1 (6.4)	41.4 (19.7)
Older[b2]					
SRD	20	11.2 (0.4)	8.75 (0.6)	97.5 (6.2)	50.0 (11.2)
Control	20	8.8 (0.7)	8.91 (0.7)	101.3 (8.9)	65.0 (21.3)
Szeszulski and Manis (*1987*)					
Medium reading[a]					
SRD	37	10.3 (1.8)	Gr 2.4 (0.7)	98.9 (10.1)	33.1
Control	14	7.1 (0.7)	Gr 2.7 (0.6)	106.0 (8.6)	48.2
High reading[b]					
SRD	15	13.2 (1.4)	Gr 4.7 (0.7)	101.1 (8.4)	54.7
Control	20	8.9 (0.9)	Gr 4.9 (0.8)	106.5 (10.9)	55.2
Vellutino and Scanlon (*1987*)					
SRD	75	12.0	Gr 3.7	104.0	59.5 (16.4)
Control	75	7.9	Gr 3.8	120.0	60.5 (18.7)

continued over

DOI: 10.1002/dys

Table 1 *continued*

Study	n	Age	Reading level	IQ	Nonwords
Baddeley, Logie, and Ellis (1988)					
SRD	15	11.9	9.1	73.0	58.0
Control	16	8.6	9.1	76.3	60.3
Holligan and Johnston (1988)					
SRD	20	8.5 (0.4)	7.0 (0.3)	103.7 (9.3)	50.9 (2.9)
Control	20	7.2 (0.3)	7.2 (0.2)	107.7 (14.9)	60.6 (13.5)
Manis, Szeszulski, Holt, and Graves (1988)					
SRD	50	9.2–14.9	PR 13.6	107.3	50.3
Control	40	6.1–10.7	PR 79.6	108.4	71.0
Siegel and Ryan (1988)					
SRD	41	—	Gr 2–6	>80	36.8
Control	76	—	Gr 2–6	>80	54.5
Olson, Wise, Conners, Rack, and Fulker (1989)					
SRD	58	15.6	11.8	>90	55.0
Control	57	10.3	11.8	>90	70.0
2. Additional studies					
Frith and Snowling (1983)					
SRD	26	10–12	8.3–10.8	100–133	36.7 (25.0)
Control	10	9–10	8.5–10.2	—	72.5 (16.7)
Conners and Olson (1990)					
SRD	115	15.6 (2.7)	6.5 (1.9)	99.6 (9.7)	−0.37 (0.99)†
Control	115	10.4 (1.6)	6.5 (1.9)	111.6 (12.2)	0.29 (0.85)
Manis, Szeszulski, Holt, and Graves (1990)					
SRD	52	11.8 (1.4)	3.9 (0.9)	108.2 (10.8)	47.8 (16.2)
Control	35	8.2 (0.7)	4.0 (0.8)	107.3 (8.5)	62.6 (16.5)
Pennington et al. (1990)					
Familial subgroup[a]					
SRD	15	25.6 (6.2)	Gr 8.5 (3.3)	96.4 (12.4)	Gr 7.0 (4.3)
Control	15	13.2 (2.8)	Gr 8.6 (3.3)	113.7 (13.0)	Gr 9.1 (4.1)
Clinic subgroup[b]					
SRD	15	30.9 (7.8)	Gr 10.5 (2.5)	98.9 (8.1)	Gr 4.8 (3.0)
Control	15	14.3 (2.8)	Gr 9.8 (3.3)	112.1 (10.4)	Gr 10.1 (3.9)
Bowey et al. (1992)					
SRD	16	9.1 (0.4)	7.9 (0.5)	104.1 (12.4)	39.0 (15.0)
Control	16	7.4 (0.2)	7.9 (0.5)	100.6 (11.2)	59.0 (12.0)
Felton and Wood (1992)					
Grade 3 subgroup[a]					
SRD	93	9.3 (0.6)	7.17	93.2 (5.6)	13.0 (11.0)
Control	93	7.1 (0.4)	7.17	100.9 (11.4)	20.0 (15.0)
Grade 5 subgroup[b]					
SRD	54	11.5 (0.5)	7.2	93.4 (8.2)	30.0 (17.7)
Control	54	7.2 (0.4)	7.25	106.6 (10.4)	40.0 (19.0)

continued over

DOI: 10.1002/dys

Table 1 *continued*

Study	*n*	Age	Reading level	IQ	Nonwords
Manis et al. (*1993*)					
SRD	21	13.0 (2.1)	Gr 4.0 (1.0)	109.6 (10.1)	48.8 (21.8)
Control	32	10.0 (1.2)	Gr 4.0 (0.6)	106.9 (8.5)	64.2 (15.9)
Bowey and Hansen (*1994*)					
SRD	20	9.3 (0.3)	7.3 (0.3)	106.3 (17.9)	30.8 (21.5)
Control	20	7.2 (1.5)	7.3 (0.3)	98.8 (11.8)	49.3 (21.9)
Murphy and Pollatsek (*1994*)					
SRD	65	11.5 (0.9)	PR 77.1 (8.8)	105.1 (10.2)	51.5 (18.9)
Control	65	7.7 (0.8)	PR 110.4 (8.4)	—	71.2 (18.8)
Vellutino et al. (*1994*)					
SRD (severe)	45	Gr 6 and 7	Gr 3.25	94.8 (9.8)	53.5 (13.0)
Control	73	Gr 2 and 3	Gr 3.08	110.9 (10.0)	49.5 (16.5)
Stothard and Hulme (*1995*)					
SRD	14	8.4 (0.6)	6.5 (0.8)	101.1 (13.5)	51.0 (19.0)
Control	14	6.7 (0.6)	6.6 (0.8)	107.6 (11.9)	71.0 (19.0)
Manis et al. (*1996*)					
SRD	51	12.4 (1.8)	4.4 (1.1)	106.3 (14.9)	61.9 (15.7)
Control	27	8.5 (0.6)	4.4 (1.2)	109.6 (12.1)	76.9 (17.5)
Snowling et al. (*1996*)					
SRD	20	10.7	8.0 (1.1)	109.1 (10.1)	47.9 (25.8)
Control	20	7.5	7.8 (0.8)	108.6 (11.7)	50.8 (28.3)
Badian (*1997*)					
SRD	28	8.8 (0.7)	PR 75.3 (6.3)	104.5 (11.3)	21.3 (16.4)
Control	24	6.9 (0.4)	PR 109.9 (12.3)	100.1 (9.0)	38.9 (18.0)
Greenberg et al. (*1997*)					
SRD	75	33.4 (11.8)	11.0 (1.3)	113.1 (16.6)	45.0 (21.0)
Control	75	10.1	10.9 (1.3)	100.6 (12.2)	68.0 (17.0)
Herrmann (*1997*)					
SRD	20	11.3 (0.8)	7.9 (0.7)	95.3 (8.4)	14.7 (6.7)
Control	20	7.8 (0.7)	8.1 (0.6)	101.5 (10.8)	22.2 (7.2)
Stanovich, Siegel and Gottardo (*1997*)					
Empirical study[a]					
SRD	68	9.0	PR 10.6	—	36.5
Control	23	7.4	PR 49.5	—	45.0
Reanalysis of Castles and Coltheart (1993) data[b]					
SRD	40	11.5	8.4	97.0	46.0
Control	17	8.5	8.7	—	69.0
Duncan and Johnston (*1999*)					
SRD	41	10.7 (0.7)	8.0 (0.8)	107.9 (13.3)	44.5 (14.3)
Control	41	7.6 (0.3)	8.2 (0.6)	110.5 (11.8)	53.5 (17.9)

[*]*d*-prime scores.
[†]Latent variable scores.
1 & 2: [a] and [b] refer to different comparisons reported within the same paper.

DOI: 10.1002/dys

Bus (1994). To achieve this, many of the procedures described in the original meta-analysis were repeated so that confounds arising from methodological differences would not be introduced. Summary statistics were retrieved from the 34 studies in one of three ways: (a) the test statistic (t, F) was explicitly reported in the paper; (b) the study provided means and standard deviations for performances on the nonword reading test, and t-statistics were computed from these data; or (c) the exact p-value was reported in the paper, or conservative estimates of significance level for differences between groups were used (no significant difference: $p = 0.50$; significant difference: $p = 0.05$). This third approach was applied to some of the studies examined by Ijzendoorn and Bus previously, where no other data were available for meta-analysis. Summary statistics from all other studies were obtained using the first two methods.

As studies included in the analysis had different sample sizes and variances, it was necessary to convert statistics onto a common scale. Repeating Ijzendoorn and Bus' (1994) methodology, the statistic used for this purpose was the Pearson Product Moment Correlation r. The r-values were derived from summary statistics using standard meta-analytic formulae (e.g. Wolf, 1985).

This method of converting summary data onto a common scale uses the pooled standard deviation for the two groups. If one group is significantly more variable in performance than the other, however, then the between group effect may not be calculated precisely. Twenty-seven comparisons reported standard deviations for group outcomes (see Table 1). In 65% of comparisons, the reading-level control groups had larger degrees of individual differences in nonword reading than SRD groups. When the difference between the two group variances was tested, however, most F-ratios were small and none was found to be statistically significant (i.e. $F < 1.0$). Therefore, the two groups demonstrated similar performance variability within studies and it was deemed appropriate to pool standard deviations to convert summary statistics into a common metric.

The standardized effect size index was the primary outcome of interest in the current research. This represents the mean performance difference in nonword reading ability between SRD and reading-level control groups, expressed in standard deviation units. The standardized difference between the means of the two groups (d) was derived from r using Cohen's (1977) formula:

$$d = \sqrt{4r^2/(1 - r^2)}$$

Finally, the effect sizes for studies were combined using the Stouffer method (1949), which weights each study by sample size. The formula for combining effect sizes is:

$$\text{Fisher } Z = \frac{\sum n \text{ Fisher } Z}{\sum n}$$

Tests of Homogeneity of Effect Sizes and Moderator Variables

Tests of homogeneity were used to explore whether the different outcomes across studies could be explained by sampling and measurement error or, alternatively,

DOI: 10.1002/dys

by studies using samples representative of different populations. The following chi-square formula was applied to test homogeneity of effect sizes:

$$\chi^2(k-1) = \sum(n-3)\ (\text{Fisher}\ Z - \text{Fisher}\ \overline{Z})^2 \quad (k = \text{number of studies})$$

To examine the reasons for heterogeneity in effect sizes, the following five moderator variables were derived from the studies:

Age: The continuous variables of mean ages of the SRD groups, of the reading-level matched control groups, and of the difference between the mean ages of the two groups, were examined. Standard deviation data were also analysed.

Reading level: The continuous variables of mean reading levels of the SRD groups, of the control groups, and of the difference between the mean reading levels of the two groups, were examined. Standard deviation data were also analysed. Most studies reported reading level in either reading age or reading grade equivalent units. These data were analysed separately because they could not be transformed onto a common measurement scale.

IQ level: The continuous variables of mean IQ scores for the SRD and control groups, and of the difference between the mean IQ scores for the two groups, were examined. Standard deviation data were also analyzed.

Type of reading test: The standardized reading tests used to match SRD and reading-level matched control groups were categorized into three groups: (i) the Word Identification subtest of the Woodcock reading batteries (Woodcock Reading Mastery Test or Woodcock Johnson Psychoeducational Battery); (ii) other word identification tests involving reading accuracy for individual words; and (iii) tests of story passage reading accuracy.

Type of intelligence test: The tests used to match SRD and control groups for level of intelligence were divided into three groups: (i) those measuring full-scale intelligence; (ii) those measuring verbal intelligence only; and (iii) those measuring nonverbal/performance intelligence only.

Data for the three continuous variables—age, reading level, and intelligence level—are presented for each study in Table 1. Table 2 shows the test instruments used.

Table 2 shows instances where two or more studies have used the same published nonword reading test as the dependent variable measure: The Word Attack subtests of the Woodcock tests; the Decoding Skills Test (DST); and the Bryant Diagnostics Test (BDT). The availability of data from testing of independent samples with the same nonword reading measure provided a unique opportunity to obtain additional information about the homogeneity of these samples. To address this, studies that collected nonword reading data using the same test instrument were grouped together, and their standard deviation data from Table 1 were compared.

These comparisons showed a narrow range of standard deviations for groups across testing occasions with the same test instrument. For example, studies that used the Word Attack subtest of the WJPB had raw score standard deviations that ranged from 2.3 to 3.7 for SRD groups, and from 2.9 to 4.2 for reading-level control groups. On this basis, it was judged that approximately the same degree of performance variability was observed on the same measure on independent testing occasions, and that groups had been sampled from a similar population of readers across studies.

DOI: 10.1002/dys

Table 2. Test instruments used in studies of the nonword reading deficit

Study	Reading test	IQ test	Nonword test
1. Original studies			
Snowling (1980)	Schonell	PPVT	exp
Snowling (1981)	Schonell	PPVT	exp (1 & 2)
Baddeley *et al.* (1982)	Schonell	WISC-R/Terman	exp
DiBenedetto *et al.* (1983)	DST-GE	PPVT	exp
Kochnower *et al.* (1983)	DST-GE	PPVT	DST
Beech and Harding (1984)	Schonell	Raven	exp
Olson *et al.* (1985)	PIAT	WISC-R	exp
Treiman and Hirsch-Pasek (1985)	WRMT	PPVT	exp
Johnston *et al.* (1987[a])	BAS	BAS	exp
Johnston *et al.* (1987[b])	BAS	BAS	exp
Szeszulski and Manis (1987[a])	Gilmore	WISC-R	exp
Szeszulski and Manis (1987[b])	Gilmore	WISC-R	exp
Vellutino and Scanlon (1987)	Gilmore	Slosson	BDT
Baddeley *et al.* (1988)	—	PIQ/Raven	exp
Manis *et al.* (1988)	WRMT	WISC-R	exp (1 & 2)
Holligan and Johnston (1988)	BAS	WISC-R	exp
Siegel and Ryan (1988)	WRAT-GE	PPVT	G-F-W SS
Olson *et al.* (1989)	PIAT	WISC-R	exp
2. Additional studies			
Frith and Snowling (1983)	BAS	WISC-R	exp
Conners and Olson (1990)	PIAT	WISC-R	exp
Manis *et al.* (1990)	WRMT-GE	WISC-R	BDT & WA
Pennington *et al.* (1990[a])	PIAT-GE	WAIS-R	WA-WJPB
Pennington *et al.* (1990[b])	PIAT-GE	WAIS-R	WA-WJPB
Bowey *et al.* (1992)	WRMT	PPVT	WA-WRMT
Felton and Wood (1992[a])	WJPB	PPVT	WA-WJPB
Felton and Wood (1992[b])	WJPB	PPVT	WA-WJPB
Manis *et al.* (1993)	WRMT	WISC-R	exp (1 & 2)
Bowey and Hansen (1994)	WRMT	PPVT	exp
Murphy and Pollatsek (1994)	WRAT-SS	WISC-R	DST
Vellutino *et al.* (1994)	Gilmore	Slosson	BDT
Stothard and Hulme (1995)	Neale	WISC-R	exp (1 & 2)
Manis *et al.* (1996)	WRMT-GE	WISC-R	exp
Snowling *et al.* (1996)	BAS	WISC-R	exp
Badian (1997)	WRMT-SS	WISC-R	WA-WRMT
Greenberg *et al.* (1997)	WRMT	PPVT	WA-WRMT
Herrmann (1997)	WRMT	WISC-III	WA-WRMT
Stanovich *et al.* (1997[a])	WRAT-PR	—	WA-WRMT
Stanovich *et al.* (1997[b])	Neale/WRMT/Gap	WISC-R	WA-WRMT
Duncan and Johnston (1999)	BAS	WISC-R	exp

BAS—British Ability Scales, Word Recognition; BDT—Bryant Decoding Skills Test; DST—Decoding Skills Test; exp—Experimental nonword reading list; one-syllable items exp (1 & 2)—Experimental nonword reading list, one- and two-syllable items; Gap—Gap Reading comprehension Test; G-F-W SS—Goldman–Fristoe–Woodcock Sound Symbols Test; Gilmore—Gilmore Oral Reading Test; Neale—Neale Analysis of Reading Ability; PIAT—Peabody Individual Achievement Test; PIQ—Performance Intelligence Quotient; PPVT—Peabody Picture Vocabulary Test-Revised; Raven—Ravens Progressive Matrices; Schonell—Schonell Graded Word Recognition Test; Slosson—Slosson Intelligence Test; WA-WJPB—Woodcock Johnston Psychoeducational Battery, Word Attack; WA-WRMT—Word Attack, Woodcock Reading Mastery Test-Revised; WAIS-R—Wechsler Adult Intelligence Scale-Revised; WISC-R/Terman—Wechsler Intelligence Scale for Children-Revised and Terman Intelligence Test; WRAT—Wide Range Achievement Test, Word Recognition; WRMT-R—Woodcock Reading Mastery Test-Revised, Word Identification.

DOI: 10.1002/dys

Procedure

Ijzendoorn and Bus' (1994) methods of retrieving summary statistics, converting them into a common metric, and combining results were used. Replication of Ijzendoorn and Bus' study was first undertaken to verify correct application of procedures. Secondly, the meta-analytic procedures were applied to the additional studies, and then to all studies to obtain the overall effect size index. Thirdly, homogeneity of effect magnitudes across studies was tested. Following this, potential moderator variables were tested to identify some of the possible causes of heterogeneity in study findings.

RESULTS

Replication of Ijzendoorn and Bus' (1994) Meta-analysis

Table 3 shows the values obtained when Ijzendoorn and Bus' (1994) data retrieval procedures and analyses were replicated using their original set of 16 studies. The d-values reported in Ijzendoorn and Bus' paper are presented in parentheses in the final column.

The effect sizes reported in Ijzendoorn and Bus' (1994) meta-analysis ranged from 0 to 1.03, with no negative effect sizes. The replication obtained a different effect size estimate for the Vellutino and Scanlon (1987) study. The effect size values derived from these data were otherwise replicated.

Replication of the overall combined effect size for group differences in nonword reading ability ($N = 1183$) yielded results similar to the previous meta-analysis: $d = 0.50$, comparable to a Fisher $Z = 0.24$, and a correlation coefficient of $r = 0.24$ (Ijzendoorn & Bus, 1994; reported $d = 0.48$; Fisher $Z = 0.24$; and $r = 0.24$). The homogeneity test of effect sizes was $\chi^2(17) = 27.09$, $p = 0.057$, which once again replicated Ijzendoorn and Bus' (1994) analysis ($\chi^2(17) = 27.94$, $p = 0.046$).

Meta-analysis of Additional Studies and All Studies

Following confirmation of the meta-analytic procedures, they were applied to the new sample of studies (Table 4). The effect sizes ranged from -0.25 to 1.58, indicating a greater range of outcomes than the first meta-analysis. The effect sizes reported in Ijzendoorn and Bus' (1994) review, however, fell within the distribution of effect sizes for the additional studies. Unlike the original meta-analysis, one study yielded results in the opposite direction than predicted (Vellutino *et al*., 1994) but group differences were not statistically significant.

When the additional studies were analysed separately ($N = 1682$), a larger combined effect size for nonword reading ability was found than obtained by Ijzendoorn and Bus (1994) previously. The overall combined effect size for these studies was $d = 0.76$, comparable to a Fisher $Z = 0.37$, and a correlation coefficient $r = 0.35$. When the effect sizes of all studies were combined, the average effect size for nonword reading ability was $d = 0.65$, equivalent to a Fisher $Z = 0.32$, and a correlation coefficient of $r = 0.31$.

Table 3. Replication of data reported in Ijzendoorn and Bus' (1994) meta-analysis

Study	Statistic	*df*	*N*	*p*	*Z*	Zfisher	*r*	r^2	*d*	(*d*)
Snowling (1980)	$t = 2.79$	52	54	0.004	2.65	0.38	0.36	0.13	0.77	(0.77)
Snowling (1981)	$F = 7.61$	38	42	0.009	2.37	0.44	0.41	0.17	0.90	(0.90)
Baddeley *et al.* (1982)	$p = 0.02$		30	0.02	2.05	0.40	0.38	0.14	0.81	(0.81)
Kochnower *et al.* (1983)	$t = 3.17$	38	40	0.015	2.17	0.50	0.46	0.21	1.03	(1.03)
DiBenedetto *et al.* (1983)	$t = 2.73$	38	40	0.005	2.58	0.44	0.41	0.17	0.90	(0.89)
Beech and Harding (1984)	$t = 1.54$	99	101	0.064	1.52	0.15	0.15	0.02	0.29	(0.31)
Olson *et al.* (1985)	$p = 0.05$		100	0.05	1.65	0.16	0.16	0.03	0.35	(0.33)
Treiman and Hirsh-Pasek (1985)	$t = 0.057$	72	74	0.48	0.05	0.01	0.01	0	0.02	(0.01)
Vellutino and Scanlon (1987)	$p = 0.433$		150	0.433	0.17	0.01	0.01	0	0.03	(0.14)
Szeszulski and Manis (1987[a])	$t = 2.69$	49	51	0.005	2.58	0.38	0.36	0.13	0.77	(0.77)
Szeszulski and Manis (1987[b])	$p = 0.50$		35	0.50	0	0	0	0	0	(0)
Johnston *et al.* (1987[a])	$t = 0.97$	38	40	0.17	0.95	0.16	0.16	0.03	0.35	(0.31)
Johnston *et al.* (1987[b])	$t = 2.79$	38	40	0.004	2.66	0.44	0.41	0.17	0.90	(0.91)
Baddeley *et al.* (1988)	$p = 0.50$		31	0.50	0	0	0	0	0	(0)
Manis *et al.* (1988)	$p = 0.05$		90	0.05	1.65	0.17	0.17	0.03	0.35	(0.35)
Holligan and Johnston (1988)	$t = 2.93$	38	40	0.002	2.88	0.48	0.45	0.20	1.00	(1.01)
Siegel and Ryan (1988)	$p = 0.0001$		110	0.0001	3.72	0.37	0.35	0.12	0.74	(0.76)
Olson *et al.* (1989)	$p = 0.001$		115	0.001	3.09	0.30	0.29	0.08	0.59	(0.60)

DOI: 10.1002/dys

Table 4. Data for additional studies on the nonword reading deficit

Study	Statistic (t)	df	N	p	Z	Zfisher	r	r^2	d
Frith and Snowling (1983)	4.06	26	28	0.0002	3.54	0.73	0.62	0.39	1.60
Conners and Olson (1990)	5.41	228	230	8.0E-08	4.80	0.35	0.34	0.11	0.72
Manis *et al*. (1990)	4.16	85	87	0.00005	3.96	0.44	0.41	0.17	0.91
Pennington *et al*. (1990[a])	1.36	28	30	0.09	1.32	0.25	0.25	0.06	0.51
Pennington *et al*. (1990[b])	4.17	28	30	0.0001	3.65	0.73	0.62	0.38	1.58
Bowey *et al*. (1992)	3.81	30	32	0.0003	3.41	0.65	0.57	0.33	1.40
Felton and Wood (1992[a])	3.54	184	186	0.0026	2.79	0.26	0.25	0.07	0.52
Felton and Wood (1992[b])	3.06	106	108	0.008	2.98	0.29	0.28	0.08	0.59
Manis *et al*. (1993)	3.04	51	53	0.007	2.88	0.87	0.39	0.17	0.85
Bowey and Hansen (1994)	2.72	38	40	0.006	2.61	0.43	0.41	0.16	0.89
Murphy and Pollatsek (1994)	5.97	128	130	1.0E-08	4.80	0.51	0.47	0.22	1.06
Vellutino *et al*. (1994)	−1.35	116	118	0.147	−1.34	−0.12	0.12	0.02	−0.25
Stothard and Hulme (1995)	1.27	24	26	0.11	1.23	0.26	0.25	0.06	0.52
Manis *et al*. (1996)	3.85	76	78	0.0001	3.67	0.42	0.40	0.16	0.88
Snowling *et al*. (1996)	0.385	38	40	0.35	0.95	0.06	0.06	0.004	0.13
Badian (1997)	3.67	50	52	0.0003	3.44	0.50	0.46	0.21	1.03
Greenberg *et al*. (1997)	7.36	142	144	6.8E-12	4.80	0.58	0.53	0.28	1.23
Herrmann (1997)	2.90	38	40	0.006	2.73	0.45	0.42	0.18	0.93
Stanovich *et al*. (1997[a])	3.61	55	57	0.0003	3.40	0.38	0.36	0.13	0.77
Stanovich *et al*. (1997[b])	2.15	89	91	0.0175	2.12	0.22	0.22	0.05	0.45
Duncan and Johnston (1999)	3.55	80	82	0.0003	3.43	0.39	0.37	0.14	0.79

DOI: 10.1002/dys

Tests of Homogeneity and Analysis of Moderator Variables

Significant heterogeneity of effect sizes was found among the additional studies, $\chi^2(20) = 39.45$, $p < 0.01$, and for the total sample, $\chi^2(38) = 79.38$, $p < 0.01$. This variability is illustrated in Figure 1, which presents the rank-ordered effect sizes and 95% confidence intervals for study comparisons.

The significant differences in effect sizes warranted analysis of moderator variables that might explain variability across studies. Data from descriptive and regression analyses involving the continuous moderator variables are presented in Table 5. These variables were analysed using the mean and standard deviation data reported for independent samples. The group difference variables were derived within studies by subtracting the reading-level control group mean from the SRD group mean value. All predictor variables were deemed statistically significant at $\alpha = 0.05$.

Investigation of age and reading level effects identified three outlier comparisons involving adult samples, which were removed. Of interest, however, two of these adult SRD and control group comparisons had large effect sizes (Greenberg *et al.*, 1997; Pennington *et al.*, 1990[b] see Tables 1 & 2). One comparison involving an adult sample did not detect a significant nonword reading group difference, but the effect size was moderate (Pennington *et al.*, 1990[a] see Tables 1 & 2).

Upon analysing age, neither the mean age of the SRD groups nor mean age of the control groups predicted study results. Similarly, neither degree of variability in the ages of the SRD samples nor the control samples predicted findings across studies. The difference in mean group ages, however, showed a weak negative relationship with effect size estimates, which was statistically significant. Studies with greater age differences between SRD and control groups tended to report smaller effect sizes (Figure 2).

To examine the effects of reading level, analyses were performed separately for reading age and reading grade data. For the studies that reported reading age data, neither the reading ages of the SRD samples nor the control samples predicted differences in effect sizes. Similarly, group differences in reading age did not moderate outcomes. Variability in reading age, however, showed an

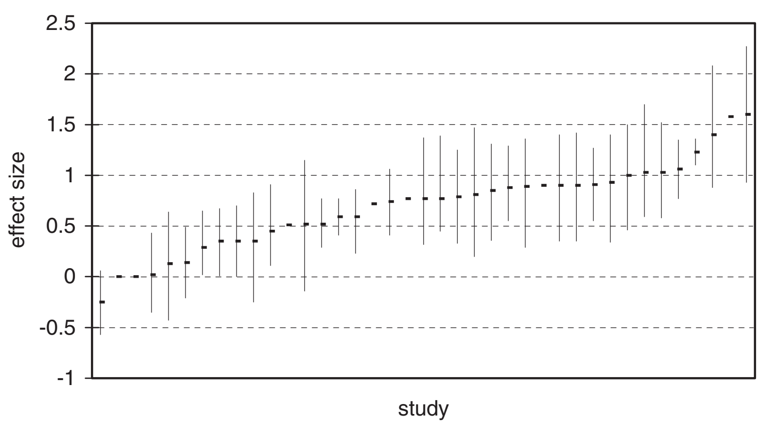

Figure 1. Effect sizes and 95% confidence intervals for group differences in nonword reading.

DOI: 10.1002/dys

Table 5. Descriptive statistics and regression analyses for moderator variables of nonword reading deficits

	Descriptive			Regression					
Variable	*M*	S.D.	Range	df	*B*	95% CI (*B*)	*t*	*p*	r^2
Chronological age (*years*)									
SRD group, *M*	11.2	1.9	8.4–15.6	31	−0.05	(−0.13, 0.03)	−1.33	0.19	0.05
RL control group, *M*	8.2	1.1	6.7–10.4	31	−0.00	(−0.15, 0.13)	−0.10	>0.50	0.00
SRD group, S.D.	1.0	0.7	0.3–2.7	18	−0.06	(−0.29, 0.18)	−0.50	>0.50	0.01
RL control group, S.D.	0.6	0.4	0.2–1.6	18	−0.18	(−0.61, 0.24)	−0.90	0.38	0.04
Group *M* difference	3.0	1.1	1.3–5.3	31	−0.14	(−0.27, −0.02)	−2.32	0.03	0.15
Reading age (*years*)									
SRD group, *M*	8.3	1.4	6.5–11.8	12	−0.02	(−0.19, 0.15)	−0.26	>0.50	0.01
RL control group, *M*	8.3	1.3	6.6–11.8	12	−0.01	(−0.19, 0.16)	−0.15	>0.50	0.00
SRD group, S.D.	0.7	0.3	0.3–1.1	8	−0.83	(−1.69, 0.03)	−2.22	0.06	0.38
RL control group, S.D.	0.6	0.3	0.2–1.3	8	−0.70	(−1.56, 0.16)	−1.88	0.10	0.31
Group *M* difference	−0.07	0.15	−0.3–0.3	12	−0.77	(−2.32, 0.79)	−1.08	0.30	0.09
Reading grade (*grade level equivalent*)									
SRD group, *M*	4.3	1.3	2.4–6.7	10	0.01	(−0.23, 0.26)	0.11	>0.50	0.00
RL control group, *M*	4.4	1.3	2.7–7.0	9	−0.01	(−0.25, 0.24)	−0.08	>0.50	0.00
SRD group, S.D.	1.1	0.4	0.7–1.9	6	0.17	(−0.81, 1.15)	0.43	>0.50	0.03
RL control group, S.D.	1.0	0.4	0.6–1.9	5	0.14	(−0.74, 1.02)	0.42	>0.50	0.03
Group *M* difference	−0.09	0.14	−0.3–0.2	9	−0.03	(−2.31, 2.25)	−0.03	>0.50	0.00
Intelligence (*standardized score*)									
SRD group, *M*	103.4	5.2	93.2–113.1	30	0.02	(−0.00, 0.04)	1.60	0.12	0.08
RL control group, *M*	106.5	5.8	94.0–120.0	28	−0.02	(−0.04, 0.01)	−1.40	0.17	0.07
SRD group, S.D.	10.9	2.9	6.2–17.9	22	0.04	(−0.02, 0.09)	1.48	0.15	0.09
RL control group, S.D.	10.5	1.9	6.4–14.9	21	0.04	(−0.05, 0.13)	0.96	0.35	0.04
Group *M* difference	−3.0	7.3	−16.0–14.0	28	0.02	(0.00, 0.04)	2.37	0.03	0.17

DOI: 10.1002/dys

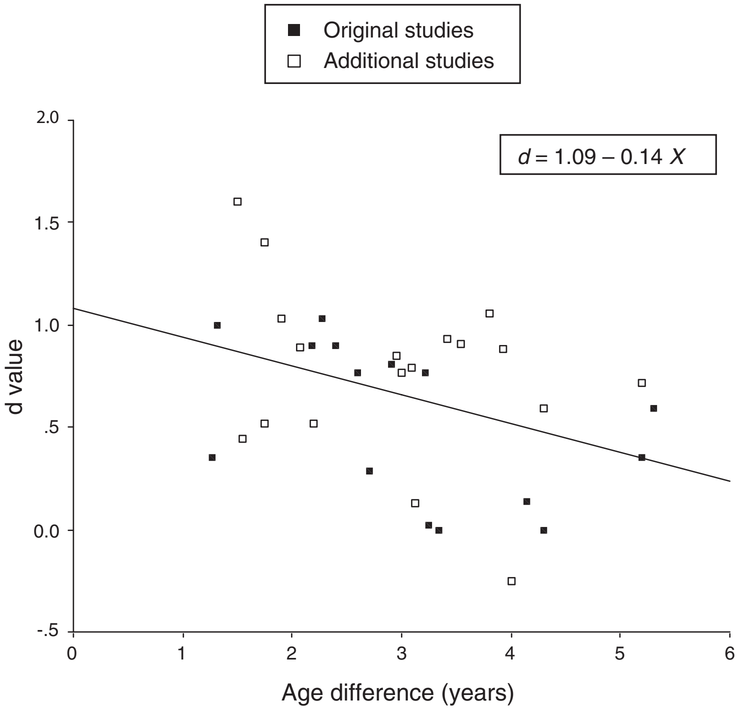

Figure 2. Scatterplot and regression line for group age differences and effect sizes.

apparently substantial and inverse relationship with effect size, suggesting a tendency for greater variability in reading ages to be associated with smaller effect sizes. These effects, however, only approached statistical significance for both SRD and control samples. Table 5 shows that the reading age standard deviation data for both groups had relatively small degrees of freedom and large confidence intervals for predictor values. No relationship was found between nonword reading outcomes and data for reading grade level: the group means, standard deviations and group mean differences in reading grades did not predict effect sizes.

The final continuous moderator variable examined was the intelligence level of groups. Neither the mean IQ scores of SRD samples nor of the control samples predicted the magnitude of effect sizes across studies. Similarly, there was no evidence for significant relationships between variability in IQ scores in the samples and study outcomes. In contrast, the difference between SRD and control groups in mean IQ scores was significantly related to effect sizes. If the SRD group scored lower than the control group in intelligence, then the effect size appeared to be smaller. The relationship between group differences in mean IQ scores and effect size estimates is shown in Figure 3.

It was noted that group differences in IQ influenced effect size estimates in the opposite direction to the findings reported by Ijzendoorn and Bus (1994). They reported that group differences in IQ and effect size correlated negatively, with smaller effect sizes associated with higher intelligence in SRD groups than control groups. Re-examination of data from their study indicated a positive, not

DOI: 10.1002/dys

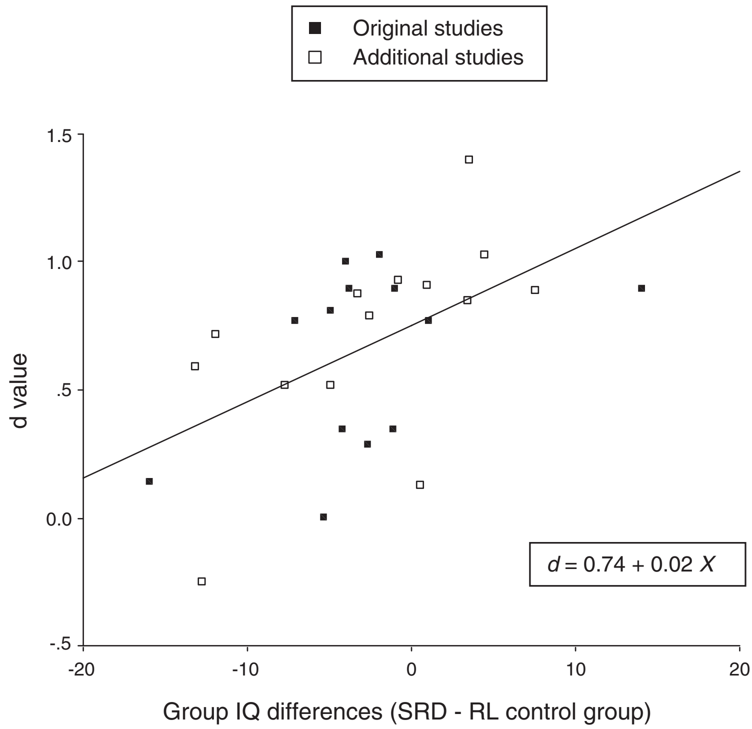

Figure 3. Scatterplot and regression line for group IQ differences and effect sizes.

a negative, relationship between these variables, as shown in Figure 3. An alternative explanation is that there is no relationship between group IQ differences and effect sizes in these studies, with the correlation portrayed being influenced by the two extreme data points. Both interpretations suggest that the result reported by Ijzendoorn and Bus was erroneous in this instance.

In the analysis of categorical moderator variables, the type of reading test used to match groups was found to influence effect sizes across studies. The group of 12 comparisons that used the Woodcock Word Identification tests were associated with the greatest mean effect size and a relatively small standard deviation of values ($d = 0.84$, S.D. = 0.27), followed by the 20 comparisons that used other measures of individual word reading accuracy ($d = 0.70$, S.D. = 0.38). Finally, the average effect size was lowest for the group of studies that used passage reading tests ($d = 0.33$, S.D. = 0.43). In fact, five of the six studies reviewed that had matched groups for text passage reading accuracy did not find evidence for nonword reading deficits.

One-way ANOVA indicated significant differences between effect sizes for levels of the reading test variable, $F(2{,}34) = 4.18$, $p = 0.02$. *Post hoc* tests (Bonferroni) showed that the effect sizes of studies that used the Woodcock Word Identification tests were significantly higher than the effect sizes of studies that used passage reading tests ($p = 0.017$). The effect sizes of studies that used measures of individual word reading accuracy other than the Woodcock test did not differ significantly from studies that used either the Woodcock test or passage

DOI: 10.1002/dys

reading tests. This might reflect the relatively large effect size standard deviations obtained for studies that used other types of individual word reading tests and passage reading tests.

Studies that used the Woodcock Word Identification subtests for reading-level group matching were more likely to use the Woodcock Word Attack subtest to measure nonword reading ability. For studies that used the Word Identification subtest, 50% used the Word Attack, compared with use of Word Attack in 17% of studies that used other individual word reading tests or passage reading tests. The effect of selecting the Word Attack subtest to measure outcomes was examined. Whereas most of the experimental nonword reading tests consisted of relatively homogeneous, monosyllabic items, the nonwords in the Word Attack test are graded according to difficulty and some items are multisyllabic. The Word Attack also might have greater reliability than other measures because it has been standardized.

The prediction tested was that use of the Word Attack subtest would be associated with larger effect sizes than use of other nonword reading tests. Across studies, the 10 comparisons that measured nonword reading with the Word Attack subtest had a higher mean effect size ($d = 0.71$) than comparisons that used other nonword reading measures ($d = 0.65$), but this difference was not statistically significant ($p > 0.05$). Similarly, when only the effect sizes of those studies that used the Woodcock Word Identification test to match groups were examined, those studies that used the Word Attack subtest to measure nonword reading (mean $d = 0.87$) did not differ significantly to studies that used other nonword reading tests (mean $d = 0.80$). The small number of comparisons was noted; however, the mean effect size for studies that used the Word Attack test were observed to be similar to those for studies that used other types of nonword reading tests. The larger effect sizes obtained for studies that used the Woodcock Word Identification tests do not appear to be explained by greater frequency of use of the Word Attack subtest.

Finally, the effect of the type of IQ test used to match groups was analysed. It was found that the average effect size was highest for the 11 comparisons that used verbal IQ measures ($d = 0.82$, S.D. $= 0.25$), followed by the 25 comparisons that used full-scale IQ measures ($d = 0.65$, S.D. $= 0.42$), and then the two studies that used nonverbal IQ measures ($d = 0.15$, S.D. $= 0.21$). Studies that used the nonverbal measures were not analysed. Independent samples t-test revealed that the mean effect size of studies that used full-scale IQ measures was not significantly different from the mean of studies that used verbal IQ tests, $t(30) = -1.47$, $p = 0.15$.

To summarize, three variables were found to have a moderating effect on study outcome: reading test type, group differences in IQ scores, and group age differences. The variance associated with the first two significant predictor variables was removed to determine whether this was sufficient to account for the heterogeneity in study outcomes. The variance associated with group age differences did not explain heterogeneity in outcomes. When d-values were adjusted for reading test type, the original variability in study outcomes was reduced substantially, but significant heterogeneity still could be detected, $\chi^2(38) = 59.19$, $p = 0.01$. Further adjustment of d-values for group differences in IQ scores, however, did reduce effect size differences to the extent that significant heterogeneity could not be detected, $\chi^2(29) = 33.5$, $p = 0.22$. Therefore, the effect

DOI: 10.1002/dys

sizes for studies were statistically homogeneous when the variance accounted for by both reading test type and group differences in IQ were removed.

There was substantial missing data, however, in the analysis involving group differences in IQ. This predictor variable may not have accounted for variance beyond that explained by reading test type in studies that specified data for both predictor variables. To test this, all cases with missing data for the two variables were excluded and the chi-square tests were repeated. In this smaller set of studies, with no adjustment significant heterogeneity was found among effect sizes, $\chi^2(29) = 59.87$, $p < 0.01$. Heterogeneity was reduced after adjustment for reading test type alone, but was statistically significant, $\chi^2(29) = 39.69$, $p = 0.07$. As reported earlier, significant variability in findings could no longer be detected when the effects of both reading test type and group IQ differences were controlled. Therefore, the combination of both reading test type and group differences in IQ accounted for much greater variability in outcomes than the reading test variable alone.

DISCUSSION

The meta-analysis involved 39 comparisons of independent groups and 2865 research participants. When data were combined from these studies, the SRD groups performed 0.65 standard deviation units below their reading-level matched peers in nonword reading on average. The results of the study were consistent with Ijzendoorn and Bus (1994) in providing quantified evidence for a nonword reading deficit in SRD. A larger combined effect size estimate, however, was obtained from the studies in the expanded database than was found previously ($d = 0.48$). Overall, the outcomes of the meta-analysis provide strong support for the phonological deficit hypothesis. The SRD groups examined were able to read nonwords, but they performed below reading-level control groups on nonword reading tasks. This deficit in phonological recoding has been argued by many to be a leading cause of reading problems (e.g. Share, 1995).

Although most SRD groups demonstrated a problem with phonological recoding, some studies obtained much larger effect sizes than others, with d-values ranging from -0.25 to 1.58. This range was not inflated artificially by a small number of outlier values. The heterogeneity in outcomes was greater than found by Ijzendoorn and Bus (1994) previously, and was statistically significant. Only a few moderator variables, however, were associated with effect sizes. Findings of the nonword reading deficit otherwise were robust to variations in many study features, and appear to be a relatively consistent feature of specific reading groups. Nonword reading deficits in SRD have been found across samples that differ in chronological age, intelligence level and reading level.

Of interest, the three significant predictor variables related to the group matching procedure. The predictor variables involved the type of test used to match groups for reading performance, and group differences in intelligence and age. These predictors had been identified in Ijzendoorn and Bus' (1994) original meta-analysis. Their effects on nonword reading outcomes were clarified further in the present study, which involved a greater number of group comparisons.

DOI: 10.1002/dys

Studies that matched groups for reading accuracy with story passages reported smaller nonword reading deficits than studies that matched groups for individual word-level reading accuracy. Empirical findings suggest that developing readers show greater reading accuracy when words are presented in a sentence or story context than when they are presented in isolation (e.g. Nicholson, 1991; Stanovich, Cunningham, & Feeman, 1984). This advantage has been explained by the use of syntactic structure to predict the words that will follow in the passage, or by the use of semantic and syntactic context to resolve ambiguity in the identities of words (Garton & Pratt, 1998). Furthermore, research findings suggest that children with SRD rely on context when reading to a greater extent than normally developing readers to compensate for their difficulties with printed word identification (Juel, 1980; Nation & Snowling, 1998).

The use of context alone is not a particularly reliable method of word identification; however, if the child is able to recode phonologically at least some of the letters in the word then the chances of reading the word correctly increase greatly (Tunmer & Chapman, 1998). Hence, children appear to use context effectively only when it is implemented with a phonological recoding strategy. It was noted, however, that five of the six studies reviewed that matched groups for text passage reading accuracy did not find evidence for the nonword reading deficit in SRD. The one study that detected a significant group difference reported that the Woodcock Word Identification test was used to assess some research participants (Castles & Coltheart, 1993; data analysis reported in Stanovich *et al.*, 1997). Therefore, both conceptually and empirically, in reading-level matched design studies it is better to use tests of individual word identification than passage reading tests when the processes involved in word-level reading are of primary interest.

Studies that used the Word Identification subtest of the Woodcock tests for reading-level matching tended to obtain large group differences in nonword reading abilities. This finding does not support claims that using the Woodcock test is associated with smaller nonword reading deficits (Ijzendoorn & Bus, 1994; Rack *et al.*, 1992). Furthermore, the larger effect sizes obtained for studies that used the Woodcock Word Identification tests do not appear to be associated with the greater frequency of use of the Word Attack subtest because the type of nonword reading test used did not moderate study outcomes. The effect sizes for studies that used the Woodcock Word Identification tests were significantly higher than those for studies that had used passage-reading tests. In Ijzendoorn and Bus' (1994) study, outcomes from passage reading tests and Woodcock Word Identification tests were examined together. The lack of available data might have necessitated use of this method previously but, clearly, it would have been inappropriate to categorize these reading tests together in the current study.

A further significant predictor variable identified was the difference between intelligence levels of the SRD and reading-level control groups. Higher intelligence levels in SRD groups relative to control groups were associated with larger nonword reading deficits. This was also found when the analysis was conducted with the original studies from Ijzendoorn and Bus' (1994) meta-analysis only. There was otherwise no relationship between intelligence level and nonword reading ability in either group. This group difference effect is difficult

DOI: 10.1002/dys

to explain, and it could be interpreted as an artefact of the group comparison. Removing the effects of this predictor, however, reduced the heterogeneity of effect sizes to a level below the sensitivity of the meta-analysis.

The difference in mean ages of SRD and reading-level control groups also predicted effect size. Larger group differences in age were associated with findings of smaller nonword reading deficits for studies involving child and adolescent SRD samples. The effect of group age differences on outcomes was relatively weak. An implication of this finding, however, is that SRD groups with severe reading problems may show smaller nonword reading deficits than groups with moderate or mild reading problems. Groups with severe problems have the greatest discrepancy between chronological age and reading level, hence, sampling of much younger control children is needed to achieve the reading-level group match.

The most likely explanation for the age difference effect is that older children with SRD have an advantage over their younger, reading-level matched peers in perceptual and cognitive maturity and educational experience, which may lead to smaller nonword reading group differences. One of the primary criticisms of the reading-level matched design is that the two samples are drawn from different populations, in which rate of reading progress is confounded with age (Jackson & Butterfield, 1989). The older participants with SRD may have acquired additional skills and strategies that mask group performance differences on nonword reading tests. Normative data from an additional control group matched for chronological age may be warranted for studies where there are relatively large differences between the chronological ages of SRD and reading-level control groups.

Other variables showed a weak relationship with nonword reading outcomes. Those studies that used verbal intelligence tests to ensure groups were of at least average intelligence, and to match groups for IQ scores, tended to report larger nonword reading deficits and smaller variability in outcomes than studies that used full-scale measures. However, the outcomes from studies that used these two types of intelligence tests were not significantly different. Those studies that used nonverbal/performance intelligence tests obtained smaller nonword reading group differences; however, this finding, based on data from two studies, may not be very reliable.

Given the relatively small number of studies ($n = 11$) that have used verbal intelligence measures, and the trend towards larger effect sizes in these studies, Rack *et al.*'s (1992) idea that individuals with adequate verbal intelligence represent genuine, unexplained cases of SRD could be explored. Further research is needed to determine whether nonword reading deficits can be detected with greater sensitivity in these individuals. Of interest, all of the studies in the meta-analysis that measured verbal intelligence used the PPVT, a receptive vocabulary test. If Rack *et al.*'s idea is pursued, it would be important to clarify whether it is adequate verbal intelligence (verbal abstract reasoning and problem solving) or oral language development (expressive and receptive language skills) that is crucial to the identification of 'unexplained' cases of SRD characterized by particular difficulties with nonword reading.

Another variable that showed a weak relationship with nonword reading outcomes was group variability in reading level. Those studies that reported larger standard deviations for reading levels of groups tended to find smaller

DOI: 10.1002/dys

nonword reading deficits. This result suggests that it is more difficult to detect nonword reading group differences when there are greater individual differences in reading skill and educational experience within samples. The effect of reading level variability explained much of the variance in nonword reading outcomes, but was found to be statistically nonsignificant and analyses were constrained by lack of data.

It is acknowledged that further research is required for verification of predictor variables identified in a meta-analysis (Wolf, 1985). Several recommendations can be made for future research, however, based on the predictor variables of nonword reading deficits identified in the meta-analysis. For reading-level matched group studies investigating skills and processes involved in printed word identification, the use of individual word-level reading tests is clearly preferable to the use of passage reading tests. Furthermore, group differences in nonword reading appear to be detected with greater sensitivity when there are smaller group differences in chronological age and less intra-group variability in reading level, although the effects of these may be relatively weak. Researchers working with SRD groups with particularly severe reading problems may need to consider the appropriateness of the reading-level matched design, evaluating the benefits of the statistical power gained from reading-level group matching against the problems associated with comparing groups sampled from different developmental populations.

The implications of the moderating effect of group differences in intelligence on nonword reading outcomes are difficult to ascertain. In the absence of an explanation for this effect, a close group match for intelligence is advised for future studies. Upon speculating on the causes of this effect, it is recognized that what we typically refer to as a 'nonword reading deficit', with the implication that nonword reading is below what would be expected, is a discrepancy between reading accuracy for real words and nonwords. The discrepancy might, however, reflect a *strength* in real word reading relative to the underlying level of phonological recoding skill. If phonological recoding skill is essential for reading development, as theories of reading behaviour propose, then an underlying phonological recoding problem clearly would constrain reading progress. By learning other reading skills and strategies during reading development, however, it would be possible to achieve higher word-level reading abilities than would be expected for the level of phonological recoding skill. Studies in which SRD groups have higher levels of intelligence than control groups may be particularly sensitive to the use of alternative strategies to compensate for phonological recoding problems and hence individuals in these studies may be relatively stronger in their real word reading.

In conclusion, the results of the meta-analysis presented in this paper indicate clearly that most SRD groups show a significant deficit in phonological recoding when measured with nonword reading tests. The magnitude of this deficit is, on average, moderate. Evidence for the nonword reading deficit was found to be robust to many variations in study design. Some variables predicted the magnitude of this deficit, however, and these related to the type of reading test used to match groups for reading level, and to between-group differences in age and intelligence level. Based on the findings of this meta-analysis, it was concluded that SRD groups demonstrate difficulties in those aspects of reading that require phonological recoding.

DOI: 10.1002/dys

ACKNOWLEDGEMENTS

This research was conducted when the first author was in receipt of an Australian Postgraduate Award scholarship.

References

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

American Psychiatric Association. (1994). *Diagnostic and statistical manual of mental disorders* (4th ed.). Washington, DC: American Psychiatric Association.

*Baddeley, A. D., Ellis, N. C., Miles, T. R., & Lewis, V. J. (1982). Developmental and acquired dyslexia: A comparison. *Cognition*, *11*, 185–196.

*Baddeley, A. D., Logie, R. H., & Ellis, N. C. (1988). Characteristics of developmental dyslexia. *Cognition*, 29, 197–228.

*Badian, N. A. (1997). Dyslexia and the double deficit hypothesis. *Annals of Dyslexia*, *47*, 69–87.

*Beech, J. R., & Harding, L. M. (1984). Phonemic processing and the poor reader from a developmental lag viewpoint. *Reading Research Quarterly*, *19*, 357–366.

*Bowey, J. A., Cain, M. T., & Ryan, S. M. (1992). A reading-level design study of phonological skills underlying fourth-grade children's word reading difficulties. *Child Development*, *62*, 999–1011.

*Bowey, J. A., & Hansen, J. (1994). The development of orthographic rimes as units of word recognition. *Journal of Experimental Child Psychology*, *58*, 465–488.

Castles, A., & Coltheart, M. (1993). Varieties of developmental dyslexia. *Cognition*, *47*, 149–180.

Cohen, J. (1977). *Statistical power analysis for the behavioral sciences*. New York: Academic Press.

Coltheart, M., Rastle, K., Ziegler, J., & Langdon, R. (2001). DRC: A dual route cascaded model of visual word recognition and reading aloud. *Psychological Review*, *108*, 204–256.

*Conners, F., & Olson, R. (1990). Reading comprehension in dyslexic and normal readers: A component skills analysis. In D. A. Balota, G. B. Flores d'Arcais, & K. Rayner (Eds.), *Comprehension processes in reading* (pp. 557–579). Hillsdale, NJ: Erlbaum.

D'Angiulli, A., & Siegel, L. S. (2003). Cognitive functioning as measured by the WISC-R: Do children with learning disabilities have distinctive patterns of performance? *Journal of Learning Disabilities*, *36*, 48–58.

*DiBenedetto, B., Richardson, E., & Kochnower, J. (1983). Vowel generation in normal and impaired readers. *Journal of Educational Psychology*, *75*, 576–582.

*Duncan, L. G., & Johnston, R. S. (1999). How does phonological awareness relate to nonword reading skills amongst poor readers?. *Reading and Writing*, *11*, 405–439.

Ehri, L. C. (1998). Grapheme-phoneme knowledge is essential for learning to read words in English. In J. L. Metsala, & L. C. Ehri (Eds.), *Word recognition and beginning literacy* (pp. 3–40). Hillsdale, NJ: Erlbaum.

*Felton, R. H., & Wood, F. B. (1992). A reading level match study of nonword reading skills in poor readers with varying IQ. *Journal of Learning Disabilities*, *25*, 318–326.

Frith, U. (1985). Beneath the surface of developmental dyslexia. In K. E. Patterson, J. C. Marshall., & M. Coltheart (Eds.), *Surface dyslexia* (pp. 301–330). Hove, East Sussex: Erlbaum.

*Frith, U., & Snowling, M. (1983). Reading for meaning and reading for sound in autistic and dyslexic children. *British Journal of Developmental Psychology*, *1*, 329–342.

Garton, A., & Pratt, C. (1998). *Learning to be literate*: *The development of spoken and written language* (2nd ed.). Oxford: Blackwell.

DOI: 10.1002/dys

*Greenberg, D., Ehri, L.C., & Perin, D. (1997). Are word-reading processes the same or different in adult literacy students and third-fifth graders matched for reading level?. *Journal of Educational Psychology, 89*, 262–275.

*Hansen, J., & Bowey, J. A. (1994). Phonological analysis skills, verbal working memory, and reading ability in second-grade children. *Child Development*, *65*, 938–950.

*Herrmann, J. A. (1997). Do children with developmental dyslexia display orthographic compensation? Unpublished Honours manuscript, La Trobe University, Bundoora, Australia.

*Holligan, C., & Johnston, R. (1988). The use of phonological information by good and poor readers in memory and reading tasks. *Memory and Cognition*, *16*, 522–532.

Ijzendoorn, M. H. V., & Bus, A. G. (1994). Meta-analytic confirmation of the nonword reading deficit in developmental dyslexia. *Reading Research Quarterly, 29*, 266–275.

Jackson, N. E., & Butterfield, E. C. (1989). Reading-level match designs: Myths and realities. *Journal of Reading Behaviour, 11*, 387–412.

*Johnston, R. J., Rugg, M. D., & Scott, T. (1987). The influence of phonology on good and poor readers when reading for meaning. *Journal of Memory and Language*, *26*, 57–68.

Juel, C. (1980). Comparison of word identification strategies with varying context, word type, and reader skill. *Reading Research Quarterly, 15*, 358–376.

*Kochnower, J., Richardson, E., & DiBenedetto, B. (1983). A comparison of the phonic decoding ability of normal and learning disabled children. *Journal of Learning Disabilities*, *16*, 348–351.

*Manis, F. R., Custodio, R., & Szeszulski, R. (1993). Development of phonological and orthographic skill: A 2-year longitudinal study of dyslexic children. *Journal of Experimental Child Psychology, 56*, 64–86.

*Manis, F. R., Seidenberg, M. S, Doi, L. M., McBride-Chang, C., & Peterson, A. (1996). On the bases of two subtypes of developmental dyslexia. *Cognition*, *58*, 157–195.

*Manis, F. R., Szeszulski, P. A., Holt, L. K., & Graves K. (1988). A developmental perspective on dyslexic subtypes. *Annals of Dyslexia*, *38*, 139–153.

*Manis, F. R., Szeszulski, P. A., Holt, L. K., & Graves, K. (1990). Variation in component word recognition and spelling skills among dyslexic children and normal readers. In T. H. Carr, & B. A. Levy (Eds.), *Reading and its development*: *Component skills approaches* (pp. 207–260). USA: Academic Press.

*Murphy, L., & Pollatsek, A. (1994). Developmental dyslexia: heterogeneity without discrete subgroups. *Annals of Dyslexia*, *44*, 120–146.

Nation, K., & Snowling, M. J. (1998). Individual differences in contextual facilitation: evidence from dyslexia and poor reading comprehension. *Child Development*, *69*, 996–1011.

Nicholson, T. (1991). Do children read words better in context or in lists? A classic study revisited. *Journal of Educational Psychology, 83*, 444–450.

*Olson, R. K., Kliegel, R., Davidson, B. J., & Foltz, G. (1985). Individual and developmental differences in reading disability. In G. E. MacKinnon, & T. G. Waller (Eds.), *Reading research*: *Advances in theory and practice* (Vol. 4, pp. 1–64). New York: Academic Press.

*Olson, R. K., Wise, B., Conners, F. A., Rack, J., & Fulker, D. (1989). Specific deficits in component reading and language skills: Genetic and environmental influences. *Journal of Learning Disabilities*, *22*, 339–348.

*Pennington, B. F., Van Orden, G. C., Smith, S. D., Green, P. A., & Haith, M. M. (1990). Phonological processing skills and deficits in adult dyslexics. *Child Development*, *61*, 1753–1778.

Plaut, D. C., McClelland, J. L., Seidenberg, M. S., & Patterson, K. E. (1996). Understanding normal and impaired word reading: Computational principles in quasi-regular domains. *Psychological Review, 103*, 56–115.

Rack, J. P., Snowling, M. J., & Olson, R. K. (1992). The nonword reading deficit in developmental dyslexia: A review. *Reading Research Quarterly, 27*, 29–53.

DOI: 10.1002/dys

Shankweiler, D., Lundquist, E., Katz, L., Stuebing, K. K., Fletcher, J. M., Brady, S., Fowler, A., Dreyer, L. G., Marchione, K. E., Shaywtiz, S. E., & Shaywitz, B. A. (1999). Comprehension and decoding: Patterns of association in children with reading difficulties. *Scientific Studies of Reading*, *3*, 69–94.

Share, D. L. (1995). Phonological recoding and self-teaching: the *sine qua non* of reading acquisition. *Cognition*, *55*, 155–218.

*Siegel, L. S., & Ryan, E. B. (1988). Development of grammatical sensitivity, phonological, and short-term memory skills in normally achieving and learning disabled children. *Developmental Psychology*, *24*, 28–37.

*Snowling, M. J. (1980). The development of grapheme-phoneme correspondence in normal and dyslexic readers. *Journal of Experimental Child Psychology*, *29*, 294–305.

*Snowling, M. J. (1981). Phonemic deficits in developmental dyslexia. *Psychological Research*, *43*, 219–234.

Snowling, M. J. (1991). Developmental reading disorders. *Journal of Child Psychology and Psychiatry*, *32*, 49–77.

Snowling, M. J. (2002). Reading development and dyslexia. In U. Goswami (Ed.), *Blackwell handbook of childhood cognitive development* (pp. 394–411). Malden, MA, USA: Blackwell Publishing.

*Snowling, M. J., Goulandris, N., & Defty, N. (1996). A longitudinal study of reading development in dyslexic children. *Journal of Educational Psychology*, *88*, 653–669.

Stanovich, K. E., Cunningham, A. E., & Feeman, D. J. (1984). Relation between early reading acquisition and word decoding with and without context: A longitudinal study of first-grade children. *Journal of Educational Psychology*, *76*, 668–677.

Stanovich, K. E., Siegel, L. S., & Gottardo, A. (1997). Converging evidence for phonological and surface subtypes of reading disability. *Journal of Educational Psychology*, *89*, 114–127.

*Stothard, S. E., & Hulme, C. H. (1995). A comparison of phonological skills in children with reading comprehension difficulties and children with decoding difficulties. *Journal of Child Psychology and Psychiatry*, *36*, 399–408.

Stouffer, S. A. (1949). *The American soldier*: *Volume 1. Adjustment during army life*. Princeton, NJ: Princeton University Press.

*Szeszulski, P. A., & Manis, F. R. (1987). A comparison of word recognition processes in dyslexic and normal readers at two reading-age levels. *Journal of Experimental Child Psychology*, *44*, 364–376.

*Treiman, R., & Hirsh-Pasek, K. (1985). Are there qualitative differences in reading behavior between dyslexic and normal readers?. *Reading Research Quarterly*, *13*, 357–364.

Tunmer, W. E., & Chapman, J. W. (1998). Language prediction skill, phonological recoding, and beginning reading. In C. Hulme, & J. R. Metsala (Eds.), *Reading and spelling*: *Development and disorders* (pp. 33–67). Mahwah, NJ: Lawrence Erlbaum.

Vellutino, F. R., Fletcher, J. M., Snowling, M. J., & Scanlon, D. M. (2004). Specific reading disabilities (dyslexia): What have we learned in the past four decades? *Journal of Child Psychology and Psychiatry*, *45*, 2–40.

*Vellutino, F. R., & Scanlon, D. M. (1987). Phonological coding, phonological awareness, and reading ability: Evidence from a longitudinal and experimental study. *Merrill-Palmer Quarterly*, *33*, 321–364.

*Vellutino, F. R., Scanlon, D. M., & Tanzman, M. S. (1994). Components of reading ability: Issues and problems in operationalizing word identification, phonological coding, and orthographic coding. In G. R. Lyon (Ed.), *Frames of reference for the assessment of learning disabilities* (pp. 279–323). Baltimore: Brooks.

Wolf, F. M. (1985). *Meta-analysis*: *Quantitative methods for research synthesis*. US: Sage.

DOI: 10.1002/dys