University of Groningen

Phonological Processing and Word Reading in Typically Developing and Reading Disabled Children

de Groot, Barry J. A.; van den Bos, Kees P.; Minnaert, Alexander E. M. G.; van der Meulen, Bieuwe F.

Published in:
Scientific Studies of Reading

DOI:
[10.1080/10888438.2014.973028](https://doi.org/10.1080/10888438.2014.973028)

IMPORTANT NOTE: You are advised to consult the publisher's version (publisher's PDF) if you wish to cite from it. Please check the document version below.

Document Version
Publisher's PDF, also known as Version of record

Publication date:
2015

[Link to publication in University of Groningen/UMCG research database](#)

Citation for published version (APA):

de Groot, B. J. A., van den Bos, K. P., Minnaert, A. E. M. G., & van der Meulen, B. F. (2015). Phonological Processing and Word Reading in Typically Developing and Reading Disabled Children: Severity Matters. *Scientific Studies of Reading*, 19(2), 166-181. <https://doi.org/10.1080/10888438.2014.973028>

Copyright

Other than for strictly personal use, it is not permitted to download or to forward/distribute the text or part of it without the consent of the author(s) and/or copyright holder(s), unless the work is under an open content license (like Creative Commons).

The publication may also be distributed here under the terms of Article 25fa of the Dutch Copyright Act, indicated by the "Taverne" license. More information can be found on the University of Groningen website: <https://www.rug.nl/library/open-access/self-archiving-pure/taverne-amendment>.

Take-down policy

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

Downloaded from the University of Groningen/UMCG research database (Pure): <http://www.rug.nl/research/portal>. For technical reasons the number of authors shown on this cover page is limited to 10 maximum.

Phonological Processing and Word Reading in Typically Developing and Reading Disabled Children: Severity Matters

Barry J. A. de Groot, Kees P. van den Bos, Alexander E. M. G. Minnaert,
and Bieuwe F. van der Meulen

University of Groningen

In this study word reading (WR) fluency was used to dichotomously classify 1,598 Dutch children at different cutoffs, indicating (very) poor or (very) good reading performance. Analysis of variance and receiver operating characteristics were used to investigate the effects of rapid automatized naming (RAN) and phonemic awareness (PA) in predicting group membership. The highest predictive values were found for the combination of RAN and PA, particularly for the poorest readers. Furthermore, results indicate that with the severity of impairment, WR is more dominated by deficient PA, which is interpreted as an enduring problem with sublexical processing. Another main result is that with the increase of reading skill, the contribution of PA diminishes, whereas the contribution of RAN remains fairly constant for the whole reading fluency continuum. These results warrant the conclusion that whereas PA hallmarks reading disability, RAN appears to be the default predictor for above-average or excellent reading proficiency.

INTRODUCTION

Reading research of the past few decades has indicated phonemic awareness (PA) and rapid automatized naming (RAN) as strong predictors of WR skill (Kirby, Desrochers, Roth, & Lai, 2008; Landerl et al., 2013; Wagner, Torgesen, Rashotte, & Pearson, 2013). An influential theoretical framework involving both processes is Wolf and Bowers's (1999) Double Deficit Hypothesis (DDH), which predicts that RAN and PA constitute more or less independent correlates of WR ability. Since the introduction of the DDH, this assumption has been affirmed on numerous occasions (e.g., Compton, DeFries, & Olson, 2001; Kirby et al., 2010; Papadopoulos, Georgiou, & Kendeou, 2009; Schatschneider, Carlson, Francis, Foorman, & Fletcher, 2002; Torppa, Georgiou, Salmi, Eklund, & Lyytinen, 2012; Torppa et al., 2013; Vukovic & Siegel, 2006). The DDH also predicts that a combination of deficient RAN and PA—a “double deficit”—is associated with the poorest level of WR performance (Wolf & Bowers, 1999, p. 424). However, as DDH studies have employed differing criteria for reading disability (RD) and have applied differing cutoffs to RAN and PA to indicate deficit levels, it remains unclear what should be considered as the poorest

level of WR as well as what is meant with a double deficit. For instance, many DDH studies have focused on dichotomous RAN and PA deficit groups, whereas no reading ability levels are distinguished (e.g., Papadopoulos et al., 2009; Torppa et al., 2012; Torppa et al., 2013). Similarly, most cross-sectional and longitudinal investigations of the predictive values of RAN and PA for reading performance (Kirby, Parrila, & Pfeiffer, 2003; Vaessen & Blomert, 2012; van den Bos & De Groot, 2012) have also regarded reading performance across-the-board.

Furthermore, the latter studies have suggested that the dynamics of RAN and PA change with age, where *alphabetic* RAN typically becomes the dominant predictor over PA, whereas the influence of the latter declines with age. However, as no (fine-grained) distinctions are made for reading ability, these reports of developmental reading-predictive trajectories typically apply to the general reading population. Of importance, there are some developmental studies that *did* include dichotomous reader group comparisons (De Jong & van der Leij, 2003; van den Bos, 2008), which seem to modify the typical pattern of RAN becoming the dominant predictor over PA in two ways. First, children with RD experience *persistent* RAN difficulties throughout primary education. Second, also *PA remains* disturbed in the RD groups, especially for more demanding tasks. These results suggest that compared to typically developing children, children with RD show different reading-predictive patterns of RAN and PA that are relatively independent of age and rather depend on reading ability level itself. Still, as these studies defined RD groups according to a *single cutoff* criterion, further differential links of RAN and PA for multiple levels of reading ability possibly remain obscured.

In general, comparative studies focusing on RAN and PA typically have employed singular and arbitrary dichotomies, such as “poor readers perform below, and typical readers above the 10th percentile” (e.g., Pennington et al., 2012; van den Bos, 2008), or “poor readers perform 1.5 standard deviations below the mean” (e.g., Brizzolara et al., 2006). Similarly, there are also plenty examples of enhanced contrasts, where a gap is created by the application of a secondary cutoff to indicate typical, or average reading performance—for instance, at the 25th percentile—to be habitually contrasted with the “below the 10th percentile” group of atypical readers (e.g., Landerl et al., 2013; Pennington et al., 2012). Thus, one study’s poor readers are not necessarily equally poor readers as in other studies. This makes the evaluation of the effects of RAN and PA across studies rather difficult. Moreover, it seems that many comparative studies employ relatively lenient criteria to indicate the “poorest” readers. This may lead to underestimation of the effects of reading-related cognitive processes in the most severely impaired groups. Finally, besides a lacking differentiation for poor reading performance, it can be noted that no previous study ever looked at different levels of *above-average* reading proficiency. However, as is made clear in the next section, similar to distinguishing more fine-grained levels of “severity”, there are reasons for hypothesizing differential involvement of RAN and PA for different reading ability levels in this group as well.

Study Aims and Hypotheses

The guiding question that emanates from the rather diffuse and restricted insight into the RAN/PA-reading links just mentioned is whether a *stratified* definition of reading (dis)ability would add to the general evidence as to how PA and RAN are related to WR. The present study addresses this question with a systematic evaluation of these relationships for different ability levels along the *whole* WR fluency continuum for a large sample of Dutch schoolchildren. Its

main objective, therefore, is to investigate the reading-predictive patterns of PA and RAN as a function of WR fluency while controlling for age.

Three hypotheses have been formulated to guide the investigations. The first hypothesis assumes that when WR fluency is more deficient, the links to RAN and PA are expected to be stronger (Hypothesis 1). This hypothesis is based on reading theory (Share, 1995) which presumes that early or poor decoding skill is primarily based on “one-to-one” phonemic and phonological correspondences (as measured by PA and RAN). Conversely, more advanced decoding ability levels are more sensitive to higher order processes such as orthographic and morphemic context, which show only weak correlations with PA and RAN (Moll, Fussenegger, Willburger, & Landerl, 2009).

Second, as PA is primarily associated with early phases of reading acquisition in which poor readers dwell (Ehri, 2005, 2013; Gough, 1996; Vellutino, Fletcher, Snowling, & Scanlon, 2004), Hypothesis 1 can be refined by the expectation of relatively stronger effects for PA relative to RAN for the lowest WR ability levels (Hypothesis 2).

Third, research has left untested how (highly) proficient reading is predicted by RAN and PA. The generally held assumption is that (highly) skilled WR is characterized by unitization, lexicalization, and automatization, which make reading a fluent process (Ehri, 2005; Landerl & Wimmer, 2000; van den Bos, 2008). Assuming that fluent WR is more strongly tapped by processes involved in automatized alphanumeric naming than by sublexical processing (Vaessen & Blomert, 2012; van den Bos & De Groot, 2012) it is expected that highly proficient reading is predominantly associated with RAN, and that the contribution of PA is diminished (Hypothesis 3).

METHOD

Participants

This study concerns a total of 1,598 Dutch children (813 female, 785 male) 7 to 14 years of age, mostly from the northern region of the Netherlands. The majority of the participants were recruited from intact classes of 13 schools for regular primary and secondary education.

In addition, the sample contains a relatively small number of children with a formal diagnosis of dyslexia, and/or either one of two other clinical conditions: specific language impairment (SLI) and attention deficit hyperactivity disorder (ADHD). ADHD (American Psychiatric Association, 2000) and SLI (Tomblin et al., 1997) are considered two of the most prevalent disorders in children. Because of the relatedness to reading of ADHD (e.g., McGrath et al., 2011; Willcutt, Pennington, Olson, Chhabildas, & Hulslander, 2005), and SLI (e.g., De Bree, Wijnen, & Gerrits, 2010; Vandewalle, Boets, Ghesquiere, & Zink, 2012), it seemed reasonable to include individuals from these groups when making general inferences about cognitive reading-related processes. The proportions of the resulting special groups are dyslexia only ($n = 44$, 2.8%), ADHD-only ($n = 15$, 0.9%), dyslexia + ADHD ($n = 17$, 1.1%), SLI only ($n = 40$, 2.5%), dyslexia + SLI ($n = 11$, 0.7%). The children with SLI were sampled from schools for special education, specializing in language difficulties. The remaining special cases attended regular schools and were added to the present sample on the basis of referral by school psychologists or external health professionals. The special case groups may be considered to consist of the most severely impaired children, and

it seems plausible that the remainder of the sample contains some undiagnosed (milder) cases as well.

Children with significant hearing loss, uncorrected vision problems, neurological disorders, or low general cognitive functioning (i.e., IQ below 80) were excluded from the study.

Data Collection

Data collection was carried out by the first author and 34 undergraduate students either at a university research facility or at the schools of the participants. For participants younger than 12 years of age, informed consent was required from their parents. Participants older than 12 were also required to give consent on their own behalf.

Reader Group Assignments

All participants were dichotomously classified at different cutoff points, or criteria, according to their WR fluency performance (see the upcoming Instruments section). Depending on the sign of the criterion, participants were classified either as RD or excellent reader (ER), on one hand, or as reference (REF), on the other hand. Eight WR criteria that differed in terms of strictness were used, with critical values varying from -2 *SD* below the mean up to 2 *SD* above the population mean (i.e., -2.0 , -1.7 , -1.5 , -1.3 , $+1.3$, $+1.5$, $+1.7$, $+2.0$). These eight conditions are referred to as *WR dichotomies*, *WR criteria*, or *WR-SD*.

Instruments

Word reading fluency. WR fluency was operationalized by averaging the performances of two standardized Dutch tests for word recognition and a standardized Dutch pseudoword reading test, measuring nonword decoding skill. All three tests are used very frequently in the Netherlands, and the reported test–retest reliabilities (see the test manuals) for the individual tests are .90 or greater. The first test (Brus & Voeten, 1999) consists of real words of varying lengths, with the raw test score being the total number of words read aloud correctly (out of 116) in 1 min. The second test uses the time needed to read 50 monosyllabic real words (subtest *Monosyl*; van den Bos & Lutje Spelberg, 2010). The pseudoword reading test (van den Bos, Lutje Spelberg, Scheepstra, & De Vries, 1994) uses the total number of items read aloud correctly (out of 116) in 2 min. According to the test manuals, the raw scores of the three tests were first converted to norm scores ($M = 10$, $SD = 3$). Next, these scores were averaged. This resulted in one composite score ($M = 10$, $SD = 3$), referred to as WR.

Rapid automatized naming. RAN was assessed by the two *alphanumeric* subtests of a frequently used Dutch standardized test battery (van den Bos & Lutje Spelberg, 2010). Participants are required to sequentially state, as quickly and accurately as possible, the names of 50 letters or digits, presented to them in columns (5×10) on a standard A4-sized paper sheet. The times to complete each subtest are converted to separate norm scores, and a composite alphanumeric RAN index, referred to as RAN_{an} ($M = 10$, $SD = 3$). The test manual reports test–retest and split-half reliabilities of .84 and .86 for the subtests, respectively.

Phonemic awareness. PA was assessed by a novel Dutch computer-assisted standardized test, measuring phonemic analysis skills (De Groot, van den Bos, & van der Meulen, 2014). The test consists of two auditory subtests—phoneme elision and phoneme substitution—spanning 2×12 trials, and three practice trials each. Item examples are “Say STREEK without the /R/” (phoneme elision) and “Exchange the first sounds (i.e., phonemes) of the first and surname KEES BOS, and state the result” (phoneme substitution). The response times and accuracies were recorded and then combined by means of a standardized scoring rule, which imposes an age-dependent time penalty when an erroneous answer is given by the participant. The values of the time penalty were estimated by an optimization procedure based on the item intercorrelations of the test. The rationale of the scoring rule is to accommodate for the undesired possibility of compensating an inaccurate performance with a very swift performance. The PA measurements comprise separate scores for the subtests (PA_{elision} , and $PA_{\text{substitution}}$), and a composite PA index score, referred to as PA_{com} ($M = 50$, $SD = 10$). The manual reports average reliabilities (lambda-2, test–retest, and parallel test) of .82, .91, and .92, respectively.

Data Preparation

To facilitate the analyses, all test scores were converted to z scores. Also, to avoid confounding of the results for RAN and PA, principal component analysis with varimax rotation and Kaiser normalization was performed on the RAN subtasks and the PA subtasks (see also Table 1). This resulted in two orthogonal components that clearly bear on RAN and PA, referred to as RANFAC and PAFAC, respectively. Eighty-five percent of total variance was explained by these components. However, as these orthogonal components do not address the commonality of RAN and PA, in addition, an encompassing composite variable, referred to as RANPA, was created by averaging RAN_{an} and PA_{com} . Thus, three variables have been analyzed—PAFAC, RANFAC, and RANPA.

Statistical Analyses

Three different statistical procedures were employed to evaluate the hypothesized reading-related effects for RANFAC, PAFAC, and RANPA. The procedures are (a) computing standardized effect

TABLE 1
Rotated Factor Loadings on RAN-Factor (RANFAC) and PA-Factor (PAFAC)

	<i>Component</i>	
	<i>RANFAC</i>	<i>PAFAC</i>
RAN_{digits}	.93	.14
RAN_{letters}	.90	.24
PA_{elision}	.26	.86
$PA_{\text{substitution}}$	.12	.91

Note. RAN = rapid automatized naming; PA = phonemic awareness.

sizes (Cohen's *d*), (b) conducting univariate and multivariate analysis of variance (MANOVA), and (c) evaluating receiver operating characteristics (ROC).

First, Cohen's *d* reveals the magnitudes of main effects for RANFAC, PAFAC, and RANPA for each reading ability level. Values of .2, .5, and .8 are considered as small, medium, or large effect sizes, respectively (Cohen, 1988). These effect sizes will be used for the evaluation of the magnitudes of hypothesized group differences. However, these effect sizes do not provide information on variance components (i.e., the unique and common contributions to reading proficiency) and on (predicted) interaction effects. For this, analysis of variance (ANOVA) was deemed appropriate.

Therefore, the second, and most central method of the present study is MANOVA, which was used to evaluate main and interaction effects for each WR dichotomy. Separate runs ($8 \times 3 = 24$) of the analyses were performed with the two orthogonal components RANFAC and PAFAC, and the standardized mean of PA_{com} and RAN_{an} (RANPA), as dependent variables, and the WR dichotomies as independent variables. This method also allowed to control for age. As the standardized scores underlying the dependent variables are based on age levels corresponding to years or semesters, the models were corrected for residual age effects by including age in months as a covariate.

To take into account obvious inequalities in terms of the reader group sizes, resulting from the application of the WR criteria, a resampling procedure comprising 1,000 iterations with the WR dichotomy as the stratification variable was performed. This procedure checks each RD or ER group against 1,000 reference groups (REF) of equal size and averages the effect sizes. To investigate the interaction effects between WR dichotomy, PAFAC and RANFAC, a similar sampling procedure was applied to procure equal numbers of children from the RD and ER groups. This downsized all groups to a subsample of 68, with the exception of the $M + 2.0SD$ ER group ($n = 30$; see the appendix) which was oversampled with a factor 2.27.

Effect sizes for the MANOVAs are presented as eta squared, with values of .01, .06, and .14 considered to be small, medium, or large effect sizes, respectively (Cohen, 1973). Note that these analyses were repeated without the resampling procedure, yielding slightly attenuated effect sizes, relative to the reported values for the resampled data.

Third, ROC analysis (e.g., Fawcett, 2006) was performed to evaluate the effectiveness of RANFAC, PAFAC, and RANPA in predicting group membership. ROC has been most widely used as a measure of the effectiveness of medical screening procedures and decision making in terms of a diagnosis (e.g., Field, Oudkerk, Pedersen, & Duffy, 2013). There are also many recent examples of educational or reading-related research that have employed ROC analysis (e.g., Georgiou, Parrila, Manolitsis, & Kirby, 2011; Landerl et al., 2013; Le Jan et al., 2011; Puolakanaho et al., 2007). It is based on standard signal detection theory with $1 - \text{specificity}$, and sensitivity interpreted as the ratio of false positives and true positives, respectively. From these ratios a so-called ROC-curve can be plotted for true positives against false positives. Together with the area under the curve (AUC), which is often taken as an index of effectiveness (e.g., Fawcett, 2006), ROC curves offer the added psychometric benefit of an insight into optimal cut-off scores for the test variables. The present study extends the traditional two-dimensional ROC analysis by adding a third dimension (i.e., eight binary WR classifiers). This yields a descriptive insight into the hypothesized dynamics of PA and RAN along the WR cutoffs, which serves the interpretation of the MANOVA results.

RESULTS

Descriptive Statistics

In Figure 1 the standardized means (z scores) of the two phonological components are plotted against the WR conditions. It is evident that the means of RAN factor (RANFAC) and the PA factor (PAFAC) show distinct patterns for the target groups across the different WR criteria. Judging from the slopes, it appears that both measures interact with the WR criteria, as their means deviate from the population mean more strongly toward the more extreme criteria. Finally, the apparent widening pattern toward the more extreme WR criteria for the RD side of the continuum (WR-SD_{RD}) is suggestive of a possible three-way factor by condition interaction, that is, RANFAC $\times$ PAFAC $\times$ WR-SD_{RD}. We return to this issue in the next section.

In the appendix, a numeric overview of the descriptive statistics for each WR dichotomy is presented, with added information about the standard deviations and the number of participants per category.

Effect Sizes

Regarding the group differences in terms of mean differences, the effect sizes (Cohen's d) of group membership on RANFAC, PAFAC, and the composite variable RANPA are plotted against the WR criteria on the horizontal axis in Figure 2.

From Figure 2 it is evident that overall effect sizes are at least substantial. There is a clear enduring influence of RAN and PA throughout the reading fluency continuum. Whether an effect is considered as positive or negative is indicated by the sign of the WR criterion (see also Figure 1). Second, RANPA yields substantial added value over the separate components in terms of predictability. Moreover, RANPA clearly is more affected in the RD groups, which

FIGURE 1 Plotted means (z -scores with standard error bars) of RANFAC and PAFAC for RD/ER versus REF groups by WR dichotomy ($-2.0sd$ through $+2.0sd$).

FIGURE 2 Standardized effect sizes of mean differences (Cohen's d with standard error bars), comparing RD/ER and REF groups for RANFAC, PAFAC, and RANPA by WR dichotomy.

seems mostly due to elevated effect sizes for PA, as opposed to a rather stable pattern for RAN across most conditions. Still, effect sizes for both components become markedly larger toward the more negative conditions, which is suggestive of an interaction with $WR-SD_{RD}$. Therefore, these results may be taken as descriptive evidence in support of Hypothesis 1.

Regarding the separate components, PAFAC seems more affected in the most severe RD group, relative to RANFAC, lending support to Hypothesis 2. Finally, perhaps the most intriguing observation about Figure 2 is the ostensive drop in impact for PAFAC when going from the negatively to the positively oriented WR criteria. Apparently, the *relative* dominance of PAFAC over RANFAC for RD children reverses into a dominance of RANFAC over PAFAC for ER children. This may be taken as descriptive evidence for Hypothesis 3.

Analysis of Variance

Table 2 summarizes the results for the MANOVAs that were performed on all eight WR dichotomies ($WR-SD_{ALL}$). Generally, these results indicate strong group predictability for both RAN and PA throughout the WR continuum, as is reflected by overall large proportions of explained variance (η^2) for each dependent variable (see A, B, & C in Table 2). In addition, there is evidence of under-additivity in terms of marginal variance explained by RANPA (see C in Table 2), as the cumulative effects of the orthogonal components RANFAC and PAFAC—particularly those of the lower bound—clearly outweigh the combined effects of RAN_{an} and PA_{com} (RANPA). Furthermore, these results indicate a quite stable pattern across all eight dichotomies for large proportions of explained variance pertaining to RANFAC, albeit somewhat more for the strictest RD criteria. In contrast, the stability found for RANFAC clearly does not apply to PAFAC. PAFAC yields large effects for RD with the largest effects found for the most strict RD criteria, whereas markedly smaller effects were found for the above-average readers.

TABLE 2
F Ratios and Effect Sizes (η^2) for RANFAC, PAFAC, and RANPA Variance Components per WR-SD

	WR-SD															
	-2.0		-1.7		-1.5		-1.3		+1.3		+1.5		+1.7		+2.0	
	<i>F</i>	η^2	<i>F</i>	η^2	<i>F</i>	η^2	<i>F</i>	η^2	<i>F</i>	η^2	<i>F</i>	η^2	<i>F</i>	η^2	<i>F</i>	η^2
RANFAC (A)	110.0	.39	130.7	.35	149.0	.29	163.4	.27	123.7	.30	89.6	.29	56.7	.28	25.5	.30
PAFAC (B)	171.7	.54	172.8	.41	186.2	.32	200.9	.32	36.7	.10	36.8	.15	24.5	.16	14.8	.20
RANPA (C)	279.1	.65	328.2	.57	360.6	.50	412.7	.48	169.1	.36	139.4	.40	84.7	.37	40.2	.40
Common (D = A+B-C)		.28		.19		.11		.11		.04		.04		.07		.10
Unique RANFAC (A-D)		.11		.16		.18		.16		.26		.25		.21		.20
Unique PAFAC (B-D)		.26		.22		.21		.21		.06		.11		.09		.10

Note. All *F* values significant with one degree of freedom and $p < .001$. RANFAC = RAN-Factor; PAFAC = PA-Factor; RANPA = average of RAN_{an} and PA_{com}; WR-SD = applied Word Reading criterion in terms of Standard Deviation from the standardized population mean.

TABLE 3
 Interaction Effects of RANFAC and PAFAC With WR-SD

	WR-SD _{RD}			WR-SD _{ER}			WR-SD _{ALL}		
	<i>F</i>	<i>p</i>	η^2	<i>F</i>	<i>p</i>	η^2	<i>F</i>	<i>p</i>	η^2
RANFAC	3.56	.015	.04	0.33	n.s.	.00	18.88	<.001	.20
PAFAC	6.35	<.001	.07	3.87	.010	.04	17.30	<.001	.18

Note. Degrees of freedom for WR-SD_{RD} and WR-SD_{ER} are $F(3, 268)$, and for WR-SD_{ALL} $F(7, 536)$. RANFAC = RAN-Factor; PAFAC = PA-Factor; WR-SD = Word Reading Dichotomy.

To evaluate the interaction effects between WR-SD_{ALL}, RANFAC, and PAFAC, three separate ANOVAs were performed on each dependent variable, with WR criterion as the grouping factor, covering (a) four levels for RD (WR-SD_{RD}), (b) four levels for ER (WR-SD_{ER}), and (c) all eight levels of the (whole) WR continuum (WR-SD_{ALL}). Table 3 summarizes the results, presenting *F* ratios, *p* values, and effect sizes (η^2) for each predictor, according to the range of included WR criteria.

Regarding the set of ANOVAs for the *RD cutoffs* (a), a significant three-way RANFAC $\times$ PAFAC $\times$ WR-SD_{RD} interaction, $F(3, 268) = 2.49$, $p = .040$, was found, indicating differential patterns for RANFAC and PAFAC as a function of WR-SD_{RD}. Decomposing this effect into two separate two-way interactions, a moderately strong RANFAC $\times$ WR-SD_{RD} interaction ($\eta^2 = .04$) and an even more substantial PAFAC $\times$ WR-SD_{RD} ($\eta^2 = .07$) interaction effect were found (see Table 3). This difference is in line with the results presented in Table 2, indicating that group membership explains accrued proportions of PAFAC variance for the strictest criteria. Indeed, as compared to the less strict WR-SD_(M-1.5SD) and WR-SD_(M-1.3SD) criteria, post hoc analysis with Tukey Honestly Significant Difference correction indicated significant PAFAC differences for WR-SD_(M-1.7SD) ($p = .056$; $p = .003$, respectively) and WR-SD_(M-2.0SD) ($p = .054$; $p = .003$, respectively) *only*. Regarding RANFAC, a similar but less strong pattern applies, with only the

contrasts with $WR-SD_{(M-1.3SD)}$ yielding (marginally) significant differences ($WR-SD_{(M-1.7SD)}$, $p = .06$; $WR-SD_{(M-2.0SD)}$, $p = .006$). In conclusion, the finding of substantial $PAFAC \times WR-SD_{RD}$ and $RANFAC \times WR-SD_{RD}$ interaction effects can be interpreted as support for Hypothesis 1. The combination of a significant three-way interaction, and, relative to $RANFAC \times WR-SD_{RD}$, a large $PAFAC \times WR-SD_{RD}$ interaction effect can be interpreted as support for Hypothesis 2.

With regard to the set of ANOVAs for the *ER cutoffs* (b), and in contrast to the previous results, $PAFAC$ clearly is less strongly (positively) affected for the above-average readers. It can be noted, however, that there is a moderately strong $PAFAC \times WR-SD_{ER}$ interaction effect ($\eta^2 = .04$; see also Table 3), which can be interpreted as a small increase of effect size for $PAFAC$ toward the more extreme $WR-SD_{ER}$ criteria. The interaction with $RANFAC$ was not significant. In line with Hypothesis 3, the finding of persisting effects for $RANFAC$ along with generally diminished effect sizes for $PAFAC$, suggests that WR is *relatively* more closely associated with RAN than with PA in the ER groups. In addition, it can be noted that, whereas the common variance components are relatively small, there is a clear difference between the *unique* variance components of $RANFAC$ and $PAFAC$, with $RANFAC$ values being considerably larger (see bottom half of Table 2).

Testing the interactions of $RANFAC$ and $PAFAC$ for the *whole continuum* (i.e., $RANFAC \times WR-SD_{ALL}$ and $PAFAC \times WR-SD_{ALL}$) (c), yielded substantial effects: $\eta^2 = .20$, and $\eta^2 = .18$, respectively (see Table 3). Together with the pattern of increasing proportions of unique variance pertaining to $RANFAC$ and diminished proportions for $PAFAC$ when going from $WR-SD_{RD}$ to $WR-SD_{ER}$ (see Table 2), these results can be interpreted as strong support for Hypothesis 3.

ROC Analyses

The previously mentioned magnitude differences and interaction effects are also made apparent by the graphical sequence depicted in Figure 3, where the ROC curves of $RANFAC$, $PAFAC$, and $RANPA$ are plotted across the WR cutoffs. The reference line (diagonal) represents pure chance, and a measurement producing a curve near or even below the diagonal can be thought of as chance performance or even worse. Conversely, the higher the AUC and the steeper the slope—implying increased sensitivity at little expense to the specificity—the more effective a measure is in predicting group membership.

Demonstrated by decreasing AUCs, Figure 3 clearly shows a pattern of waning $PAFAC$ predictability as WR skill increases, whereas $RANFAC$ yields a much more stable pattern. More generally, going from negatively to positively oriented dichotomies, the reversal in terms of the components' predictive values is obvious. These results may be taken as additional support for Hypothesis 3. This also follows from the high AUCs, sensitivities, and specificities. Table 4 provides an overview of the AUCs. The highest AUC value is found for $RANPA$ (.96) at the strictest RD criterion ($WR-SD_{M-2.0SD}$). If one opts for a 95% accurate reader group prediction (accuracy = true positives + true negatives/ N ; cf. Fawcett, 2006), the optimal cutoff score for $RANPA$ is $M - 1.5 SD$. This cutoff yields a sensitivity of .77 and a specificity of .95. As a final remark, it can be noted that, for this criterion the $PAFAC$ -curve approximates the $RANPA$ -curve, indicating a stronger impact for $PAFAC$ relative to $RANFAC$. Together, these results may be taken as additional support for Hypothesis 2.

FIGURE 3 ROC curves for RANFAC, PAFAC, and RANPA by WR criterion.

TABLE 4
Areas Under the Curve for RANFAC, PAFAC, and RANPA per WR-SD

	WR-SD							
	−2.0	−1.7	−1.5	−1.3	+1.3	+1.5	+1.7	+2.0
RANFAC	.86	.84	.82	.80	.82	.81	.81	.83
PAFAC	.92	.87	.83	.82	.68	.71	.72	.76
RANPA	.96	.94	.92	.91	.87	.89	.88	.89

Note. RANFAC = RAN-Factor; PAFAC = PA-Factor; RANPA = encompassing composite variable of RAN_{an} and PA_{com}; WR-SD = applied Word Reading criterion in terms of Standard Deviation from the standardized population mean.

DISCUSSION

The present study investigated whether RAN and PA are *differentially* linked to WR for various severity levels of RD, as well as for various degrees of excellency in above-average readers. This systematic evaluation was motivated by two major concerns, the first one being comparability of previous results and the second one the hitherto neglected question as to what can be learned from excellence.

Addressing these issues, the present approach of discerning multiple severity levels of RD within a single large sample of readers has provided a comprehensive and augmented insight into the dependencies of the RAN/PA-reading relationships on the choice of WR criterion used to define RD. A main result is that *severity matters*. Although all poor readers—across all age levels investigated—showed marked problems in PA and RAN, the present study indicates that the effects of reader group membership are much stronger for—to take our most extreme case—the “WR-SD_(M−2.0SD)” criterion than for the commonly used RD criterion of “WR-SD_(M−1.3SD).” Moreover, we found increasingly larger effects for the strictest RD criteria, that is, WR-SD_(M−1.7SD) and WR-SD_(M−2.0SD), as compared to the more lenient ones. These results are in line with the prediction that a “double deficit” is present among the poorest readers (Papadopoulos et al., 2009; Wolf & Bowers, 1999) and, what is more, suggest that the Double Deficit Hypothesis gains momentum if more severely impaired readers are considered. A further specification of these findings is that the effects for PA seem consistently larger than for RAN, with—conform the prediction (Hypothesis 2)—differences getting larger with the severity of RD. This finding, thus, suggests that PA is a more effective cognitive marker for the poorest readers.

The second motive for the present study and the adopted approach of differentiating for reading ability levels was to include the virtually forgotten group of highly proficient readers. In terms of theoretical underpinnings as well as treatment goals of RD, it seems important to ask what can be learned from those who clearly have mastered decoding. In line with general reading theory (De Jong & van der Leij, 2003; Ehri, 2005; Gough, 1996; van den Bos, 2008) it was hypothesized that highly proficient reading, as opposed to deficient reading, is characterized by a relatively strong link to highly automatized access to alphanumeric networks (as measured by RAN_{an}) and a decreased necessity of sublexical analytical processes (as measured by PA). We have found strong evidence to support this hypothesis, as our data indicate a striking reversal in terms of RAN and PA deviation means (Figure 1), as well as effect sizes (Figures 2 and 3), when switching

from RD to above-average reader groups. It should be noted, however, that the dominance of RAN over PA in above-average reader groups is *relative* and due to diminished effect sizes of sublexical processing in the presence of a rather stable contribution of RAN across WR ability levels. Thus, in above-average readers, RAN seems to be the dominant predictor *by default*. The finding of RAN being a rather stable correlate of reading fluency across ability levels is a new one, at least in comparison to previous nondevelopmental, and developmental studies for the general population. For example, in their meta-study Swanson, Trainin, Necochea, and Hammill (2003) reported that RAN-Reading correlations varied only minimally with age but that these correlations were weaker in poor readers than in skilled readers. Other studies, however, report increasing correlations with age (Vaessen & Blomert, 2010; van den Bos & De Groot, 2012). Of interest, our results indicate neither an increase with age nor a decrease with impairment. For future studies, it would, therefore, be quite informative to cross-validate the present findings by means of a comprehensive developmental design that discerns reading ability levels as well.

In conclusion, the presently adopted “multiple word reading cutoff” approach to studying reading-related cognitive processes offers a comprehensive alternative to selective dichotomous group comparisons, and unitary age-based developmental approaches. With the provision of large to very large effect sizes for all reading proficiency levels, while cutting across age levels, our results are in line with the general consensus on the pivotal and enduring importance of RAN and PA to reading (dis)abilities (Kirby et al., 2008; Landerl et al., 2013; Logan, Schatschneider, & Wagner, 2011; Torppa et al., 2012; van den Bos, 2008). However, it should be noted that larger effects of both predictors were established for the most severely impaired readers. Furthermore, our results do indicate that the general involvement of phonological processing decreases with WR fluency approaching normal levels. This general pattern makes sense if assuming that advanced reading ability levels involve more “higher order” variance, for example, pertaining to orthographic processing and linguistic proficiency (Share, 1995). However, this not to deny that RAN clearly constitutes a stable and significant cognitive correlate across all reading proficiency levels, whereas sublexical processing, that is, PA, seems to be far more specific to the RD groups. Therefore, PA generally appears to be the better candidate to hallmark RD.

ACKNOWLEDGMENTS

We are grateful to all children, teachers, and university students who participated in the study. We are also greatly indebted to the three anonymous reviewers and the associate editor Dr. Karin Landerl for their helpful comments and suggestions.

REFERENCES

- American Psychiatric Association. (2000). *Diagnostic and statistical manual of mental disorders* (4th ed., text rev.). Washington, DC: Author.
- Brizzolara, D., Chilosi, A., Cipriani, P., di Filippo, G., Gasperini, F., Mazzotti, S., . . . Zoccolotti, P. (2006). Do phonologic and rapid automatized naming deficits differentially affect dyslexic children with and without a history of language delay? A study of Italian dyslexic children. *Cognitive and Behavioral Neurology*, 19, 141–149. doi:10.1097/01.wnn.0000213902.59827.19
- Brus, B. T., & Voeten, M. J. M. (1999). *Een minuut test* [One minute test]. Amsterdam, The Netherlands: Pearson.

- Cohen, J. (1973). Eta-squared and partial eta-squared in communication science. *Human Communication Research*, 28, 473–490.
- Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Hillsdale, NJ: Erlbaum.
- Compton, D. L., DeFries, J. C., & Olson, R. K. (2001). Are RAN- and phonological awareness-deficits additive in children with reading disabilities? *Dyslexia*, 7, 125–149. doi:[10.1002/dys.198](https://doi.org/10.1002/dys.198)
- De Bree, E., Wijnen, F., & Gerrits, E. (2010). Non-word repetition and literacy in Dutch children at-risk of dyslexia and children with SLI: Results of the follow-up study. *Dyslexia*, 16, 36–44. doi:[10.1002/dys.395](https://doi.org/10.1002/dys.395)
- De Groot, B. J. A., van den Bos, K. P., & van der Meulen, B. F. (2014). *Fonemische Analyse Test Revised (FAT-R)* [Phonemic Analysis Test Revised]. Amsterdam, The Netherlands: Pearson.
- De Jong, P. F., & van der Leij, A. (2003). Developmental changes in the manifestation of a phonological deficit in dyslexic children learning to read a regular orthography. *Journal of Educational Psychology*, 95, 22–40. doi:[10.1037/0022-0663.95.1.22](https://doi.org/10.1037/0022-0663.95.1.22)
- Ehri, L. C. (2005). Learning to read words: Theory, findings, and issues. *Scientific Studies of Reading*, 9, 167–188. doi:[10.1207/s1532799xssr0902_4](https://doi.org/10.1207/s1532799xssr0902_4)
- Ehri, L. C. (2013). Orthographic mapping in the acquisition of sight word reading, spelling memory, and vocabulary learning. *Scientific Studies of Reading*, 18, 5–21. doi:[10.1080/10888438.2013.819356](https://doi.org/10.1080/10888438.2013.819356)
- Fawcett, T. (2006). An introduction to ROC analysis. *Pattern Recognition Letters*, 27, 861–874. doi:[10.1016/j.patrec.2005.10.010](https://doi.org/10.1016/j.patrec.2005.10.010)
- Field, J. K., Oudkerk, M., Pedersen, J. H., & Duffy, S. W. (2013). Prospects for population screening and diagnosis of lung cancer. *The Lancet*, 382, 732–741. doi:[10.1016/S0140-6736\(13\)61614-1](https://doi.org/10.1016/S0140-6736(13)61614-1)
- Georgiou, G. K., Parrila, R., Manolitsis, G., & Kirby, J. R. (2011). Examining the Importance of Assessing Rapid Automatized Naming (RAN) for the Identification of Children with Reading Difficulties. *Learning Disabilities—A Contemporary Journal*, 9, 5–26.
- Gough, P. B. (1996). How children learn to read and why they fail. *Annals of Dyslexia*, 46, 1–20. doi:[10.1007/BF02648168](https://doi.org/10.1007/BF02648168)
- Kirby, J. R., Desrochers, A., Roth, L., & Lai, S. S. V. (2008). Longitudinal predictors of word reading development. *Literacy Development in Canada*, 49, 103–110.
- Kirby, J. R., Georgiou, G. K., Martinussen, R., Parrila, R., Bowers, P., & Landerl, K. (2010). Naming speed and reading: From prediction to instruction. *Reading Research Quarterly*, 45, 341–362. doi:[10.2307/27822891](https://doi.org/10.2307/27822891)
- Kirby, J. R., Parrila, R. K., & Pfeiffer, S. L. (2003). Naming speed and phonological awareness as predictors of reading development. *Journal of Educational Psychology*, 95, 453–464. doi:[10.1037/0022-0663.95.3.453](https://doi.org/10.1037/0022-0663.95.3.453)
- Landerl, K., Ramus, F., Moll, K., Lyytinen, H., Leppänen, P. H. T., Lohvansuu, K., . . . Schulte-Körne, G. (2013). Predictors of developmental dyslexia in European orthographies with varying complexity. *Journal of Child Psychology and Psychiatry*, 54, 686–694. doi:[10.1111/jcpp.12029](https://doi.org/10.1111/jcpp.12029)
- Landerl, K., & Wimmer, H. (2000). Deficits in phoneme segmentation are not the core problem of dyslexia: Evidence from German and English children. *Applied Psycholinguistics*, 21, 243–262.
- Le Jan, G., Le Bouquin-Jeannès, R., Costet, N., Trolès, N., Scalart, P., Pichancourt, D., . . . Gombert, J.-E. (2011). Multivariate predictive model for dyslexia diagnosis. *Annals of Dyslexia*, 61, 1–20. doi:[10.1007/s11881-010-0038-5](https://doi.org/10.1007/s11881-010-0038-5)
- Logan, J. A. R., Schatschneider, C., & Wagner, R. K. (2011). Rapid serial naming and reading ability [electronic resource] the role of lexical access. *Reading and Writing*, 24(1).
- McGrath, L. M., Pennington, B. F., Shanahan, M. A., Santerre-Lemmon, L. E., Barnard, H. D., Willcutt, E. G., . . . Olson, R. K. (2011). A multiple deficit model of reading disability and attention-deficit/hyperactivity disorder: Searching for shared cognitive deficits. *Journal of Child Psychology and Psychiatry*, 52, 547–557.
- Moll, K., Fussenegger, B., Willburger, E., & Landerl, K. (2009). RAN is not a measure of orthographic processing. Evidence From the asymmetric German orthography. *Scientific Studies of Reading*, 13, 1–25. doi:[10.1080/10888430802631684](https://doi.org/10.1080/10888430802631684)
- Papadopoulos, T. C., Georgiou, G. K., & Kendeou, P. (2009). Investigating the double-deficit hypothesis in Greek: Findings from a longitudinal study. *Journal of Learning Disabilities*, 42, 528–547. doi:[10.1177/0022219409338745](https://doi.org/10.1177/0022219409338745)
- Pennington, B. F., Santerre-Lemmon, L., Rosenberg, J., MacDonald, B., Boada, R., Friend, A., . . . Olson, R. K. (2012). Individual prediction of dyslexia by single versus multiple deficit models. *Journal of Abnormal Psychology*, 121, 212–224.
- Puolakanaho, A., Ahonen, T., Aro, M., Eklund, K., Leppänen, P. H. T., Poikkeus, A.-M., . . . Lyytinen, H. (2007). Very early phonological and language skills: Estimating individual risk of reading disability. *Journal of Child Psychology & Psychiatry*, 48, 923–931. doi:[10.1111/j.1469-7610.2007.01763.x](https://doi.org/10.1111/j.1469-7610.2007.01763.x)

- Schatschneider, C., Carlson, C. D., Francis, D. J., Foorman, B. R., & Fletcher, J. M. (2002). Relationship of rapid automatized naming and phonological awareness in early reading development: Implications for the double-deficit hypothesis. *Journal of Learning Disabilities*, 35, 245–256. doi:[10.1177/002221940203500306](https://doi.org/10.1177/002221940203500306)
- Share, D. L. (1995). Phonological recoding and self-teaching: Sine qua non of reading acquisition. *Cognition*, 55, 151–218. doi:[10.1016/0010-0277\(94\)00645-2](https://doi.org/10.1016/0010-0277(94)00645-2)
- Swanson, H. L., Trainin, G., Necoechea, D. M., & Hammill, D. D. (2003). Rapid naming, phonological awareness, and reading: A meta-analysis of the correlation evidence. *Review of Educational Research*, 73, 407–440.
- Tomblin, J. B., Records, N. L., Buckwalter, P., Zhang, X., Smith, E., & O'Brien, M. (1997). Prevalence of specific language impairment in kindergarten children. *Journal of Speech, Language, and Hearing Research*, 40, 1245–1260.
- Torppa, M., Georgiou, G., Salmi, P., Eklund, K., & Lyytinen, H. (2012). Examining the double-deficit hypothesis in an orthographically consistent language. *Scientific Studies of Reading*, 16, 287–315. doi:[10.1080/10888438.2011.554470](https://doi.org/10.1080/10888438.2011.554470)
- Torppa, M., Parrila, R., Niemi, P., Lerkkanen, M.-K., Poikkeus, A.-M., & Nurmi, J.-E. (2013). The double deficit hypothesis in the transparent Finnish orthography: A longitudinal study from kindergarten to Grade 2. *Reading and Writing*, 26, 1353–1380. doi:[10.1007/s11145-012-9423-2](https://doi.org/10.1007/s11145-012-9423-2)
- Vaessen, A., & Blomert, L. (2010). Long-term cognitive dynamics of fluent reading development. *Journal of Experimental Child Psychology*, 105, 213–231.
- Vaessen, A., & Blomert, L. (2012). The cognitive linkage and divergence of spelling and reading development. *Scientific Studies of Reading*, 17, 89–107. doi:[10.1080/10888438.2011.614665](https://doi.org/10.1080/10888438.2011.614665)
- van den Bos, K. P. (2008). Word-reading development, the double-deficit hypothesis, and the diagnosis of dyslexia. *Educational and Child Psychology*, 25, 51–69.
- van den Bos, K. P., & De Groot, B. J. A. (2012, July). *Development of word-reading speed differentially linked to the development of phonemic analysis skills and alphanumeric naming speed*. Poster session presented at the annual meeting of SSSR, Montreal, Canada.
- van den Bos, K. P., & Lutje Spelberg, H. C. (2010). *Continu Benoemen en Woorden Lezen (CB&WL)* [Test for Continuous Naming and Word Reading]. Amsterdam, The Netherlands: Boom.
- van den Bos, K. P., Lutje Spelberg, H. C., Scheepstra, A. J. M., & De Vries, J. R. (1994). *De Klepel [The Klepel, a test for pseudoword reading]*. Amsterdam, The Netherlands: Pearson.
- Vandewalle, E., Boets, B., Ghesquiere, P., & Zink, I. (2012). Development of phonological processing skills in children with specific language impairment with and without literacy delay: A 3-year longitudinal study. *Journal of Speech, Language, and Hearing Research*, 55, 1053–1067.
- Vellutino, F. R., Fletcher, J. M., Snowling, M. J., & Scanlon, D. M. (2004). Specific reading disability (dyslexia): What have we learned in the past four decades? *Journal of Child Psychology and Psychiatry*, 45, 2–40.
- Vukovic, R. K., & Siegel, L. S. (2006). The double-deficit hypothesis: A comprehensive analysis of the evidence. *Journal of Learning Disabilities*, 39, 25–47. doi:[10.1177/00222194060390010401](https://doi.org/10.1177/00222194060390010401)
- Wagner, R. K., Torgesen, J. K., Rashotte, C. A., & Pearson, N. A. (2013). *CTOPP 2: Comprehensive test of phonological processing*. Austin, TX: PRO-ED.
- Willcutt, E. G., Pennington, B. F., Olson, R. K., Chhabildas, N., & Hulslander, J. (2005). Neuropsychological analyses of comorbidity between reading disability and attention deficit hyperactivity disorder: In search of the common deficit. *Developmental Neuropsychology*, 27, 35–78.
- Wolf, M., & Bowers, P. G. (1999). The double-deficit hypothesis for the developmental dyslexias. *Journal of Educational Psychology*, 91, 415–438.

APPENDIX

TABLE A1 shows the mean test scores (z scores) and their standard deviations (between parentheses) for the following variable groups: word reading (WR), phonemic awareness {i.e., PA_{elision} , $PA_{\text{substitution}}$, the index variable PA_{com} (elision & substitution), and the orthogonal factor PAFAC}, rapid automatized naming {i.e., RAN_{digits} , RAN_{letters} , the alphanumeric index RAN_{an} (digits & letters), and the orthogonal factor RANFAC}, and the composite variable RANPA (RAN_{an} & PA_{com}).

<i>WR-SD</i>	<i>Group</i>	<i>WR</i>	<i>PA_{elision}</i>	<i>PA_{substitution}</i>	<i>RAN_{digits}</i>	<i>RAN_{letters}</i>	<i>PA_{com}</i>	<i>RAN_{an}</i>	<i>RANFAC</i>	<i>PAFAC</i>	<i>RANPA</i>
-2.0	REF	0.09	0.05	-0.03	0.01	-0.06	-0.01	-0.02	0.04	0.05	0.11
	n = 1,519	(0.94)	(1.00)	(1.05)	(1.06)	(1.04)	(1.05)	(1.01)	(0.95)	(0.94)	(0.88)
	RD	-2.47	-1.83	-1.98	-1.45	-1.72	-2.12	-1.61	-1.33	-1.82	-2.12
	n = 79	(0.33)	(1.03)	(1.03)	(0.77)	(0.89)	(0.92)	(0.75)	(0.67)	(0.84)	(0.79)
-1.7	REF	0.13	0.07	0.00	0.04	-0.03	0.02	0.01	0.07	0.09	0.16
	n = 1,485	(0.90)	(0.99)	(1.03)	(1.05)	(1.02)	(1.03)	(1.00)	(0.97)	(0.91)	(0.84)
	RD	-2.29	-1.57	-1.73	-1.36	-1.62	-1.85	-1.52	-1.16	-1.53	-1.82
	n = 113	(0.39)	(1.13)	(1.18)	(0.80)	(0.87)	(1.10)	(0.76)	(0.74)	(1.05)	(0.89)
-1.5	REF	0.19	0.10	0.03	0.07	0.00	0.05	0.04	0.10	0.08	0.20
	n = 1,440	(0.86)	(0.97)	(1.01)	(1.04)	(1.02)	(1.01)	(0.99)	(0.96)	(0.92)	(0.82)
	RD	-2.09	-1.36	-1.48	-1.20	-1.42	-1.60	-1.34	-1.06	-1.33	-1.60
	n = 158	(0.46)	(1.19)	(1.23)	(0.88)	(0.89)	(1.17)	(0.81)	(0.80)	(1.13)	(0.94)
-1.3	REF	0.23	0.13	0.06	0.09	0.03	0.08	0.07	0.12	0.11	0.23
	n = 1,397	(0.83)	(0.96)	(0.99)	(1.04)	(1.01)	(0.99)	(0.99)	(0.96)	(0.91)	(0.79)
	RD	-1.94	-1.23	-1.38	-1.09	-1.31	-1.47	-1.23	-0.97	-1.21	-1.45
	n = 201	(0.50)	(1.16)	(1.23)	(0.88)	(0.90)	(1.16)	(0.81)	(0.82)	(1.12)	(0.94)
+1.3	REF	-0.22	-0.12	-0.19	-0.17	-0.26	-0.20	-0.22	-0.13	-0.10	-0.11
	n = 1,449	(0.95)	(1.07)	(1.14)	(1.05)	(1.05)	(1.13)	(1.01)	(0.98)	(1.03)	(0.96)
	ER	1.76	0.72	0.56	1.03	1.00	0.71	1.05	1.02	0.52	1.12
	n = 149	(0.36)	(0.84)	(0.73)	(0.82)	(0.82)	(0.80)	(0.76)	(0.77)	(0.67)	(0.59)
+1.5	REF	-0.18	-0.11	-0.18	-0.15	-0.23	-0.18	-0.19	-0.11	-0.09	-0.09
	n = 1,488	(0.97)	(1.07)	(1.14)	(1.06)	(1.05)	(1.13)	(1.02)	(0.98)	(1.02)	(0.97)
	ER	1.88	0.85	0.68	1.12	1.07	0.86	1.13	1.06	0.61	1.20
	n = 110	(0.33)	(0.80)	(0.57)	(0.82)	(0.83)	(0.69)	(0.76)	(0.81)	(0.62)	(0.58)
+1.7	REF	-0.13	-0.08	-0.16	-0.12	-0.20	-0.16	-0.16	-0.08	-0.07	-0.06
	n = 1,528	(1.00)	(1.08)	(1.13)	(1.07)	(1.07)	(1.13)	(1.03)	(0.99)	(1.02)	(0.98)
	ER	2.04	0.83	0.76	1.17	1.11	0.89	1.19	1.09	0.63	1.24
	n = 70	(0.32)	(0.88)	(0.49)	(0.87)	(0.86)	(0.71)	(0.80)	(0.83)	(0.61)	(0.64)
+2.0	REF	-0.08	-0.06	-0.14	-0.09	-0.17	-0.13	-0.12	-0.05	-0.06	-0.03
	n = 1,568	(1.03)	(1.08)	(1.13)	(1.08)	(1.08)	(1.13)	(1.04)	(1.00)	(1.02)	(0.99)
	ER	2.30	0.97	0.92	1.24	1.24	1.03	1.30	1.17	0.81	1.42
	n = 30	(0.32)	(1.04)	(0.55)	(0.87)	(0.88)	(0.84)	(0.82)	(0.83)	(0.66)	(0.73)

Note. Results are grouped according to eight dichotomous word reading cutoffs (WR-SD), with the number of participants (n) indicated for each reference (REF) or focus (RD/ER) group.