Brandenburg, Janin; Klesczewski, Julia; Schuchardt, Kirsten; Fischbach, Anne; Büttner, Gerhard; Hasselhorn, Marcus

Phonological processing in children with specific reading disorder versus typical learners. Factor structure and measurement invariance in a transparent orthography

formal und inhaltlich überarbeitete Version der Originalveröffentlichung in:
formally and content revised edition of the original source in:
The Journal of educational psychology 109 (2017) 5, S. 709-726

10.1037/edu0000162

Phonological Processing in Children with Specific Reading Disorder versus Typical Learners: Factor Structure and Measurement Invariance in a Transparent Orthography

Janin Brandenburg[1,2], Julia Klesczewski[2,3], Kirsten Schuchardt[4], Anne Fischbach[1,2], Gerhard Büttner[2,3], and Marcus Hasselhorn[1,2,3]

1) German Institute for International Educational Research (DIPF), Schloßstraße 29, 60486 Frankfurt am Main, Germany
2) Center for Research on Individual Development and Adaptive Education of Children at Risk (IDeA), Schloßstraße 29, 60486 Frankfurt am Main, Germany
3) Goethe-University Frankfurt, Institute of Psychology, Theodor-W.-Adorno-Platz 6, 60629 Frankfurt am Main, Germany
4) University of Hildesheim, Universitätsplatz 1, 31141 Hildesheim, Germany

This study was part of the longitudinal research project RAVEN (Differential Diagnostic Relevance of Working Memory in Children with Learning Disorders), exploring the developmental interplay between working memory and school achievement in children with learning disabilities. RAVEN is a multi-centric study with data collection carried out in three federal states of Germany. The authors thank Christina Balke-Melcher and Claudia Mähler of the University of Hildesheim (Germany) as well as Dietmar Grube and Claudia Schmidt of the University of Oldenburg (Germany) for their involvement in carrying out this study.

This research is part of the research initiative *Developmental Disorders of Scholastic Skills*, funded by the German Federal Ministry of Education and Research (BMBF; Grant

Number: 01GJ1012A-D). The study was additionally funded by the IDeA Center, Frankfurt/Main. Correspondence concerning this article should be addressed to Janin Brandenburg, DIPF, Schloßstraße 29, 60486 Frankfurt am Main, Germany. E-mail: brandenburg@dipf.de.

Phonological Processing in Children with Specific Reading Disorder versus Typical Learners:

Factor Structure and Measurement Invariance in a Transparent Orthography

Revision submission date: August 26, 2016

Abstract

Although children with specific reading disorder (RD) have often been compared to typically achieving children on various phonological processing (PP) tasks, to our knowledge no study so far has examined whether the structure of PP applies to both groups of children alike. According to Wagner and Torgesen (1987), PP consists of three distinct constructs: phonological awareness (PA), rapid automatized naming (RAN), and the phonological loop (PL) of working memory. The present study examined whether this PP model which was originally developed for English orthography is also applicable to a more transparent language such as German. Furthermore, we tested whether the structure of PP is invariant across typically achieving children and children with RD. Therefore, 209 German-speaking third graders (100 typical learners and 109 children with RD) completed a comprehensive test battery assessing PA, RAN, and PL. Using confirmatory factor analyses, we compared the latent structure of these PP skills across both groups. The study yielded three important findings: First, Wagner and Torgesen's (1987) model transfers to the German language and its orthography with transparent grapheme-to-phoneme correspondences. Second, the tripartite structure of PP was evident across both groups (factorial invariance). Third, group invariance was also found for the measurement and structural components of the model (measurement invariance). These findings suggest that the nature of PP is invariant across typically achieving children and children with RD acquiring the transparent orthography of German. Theoretical and practical implications are discussed.

Keywords: specific reading disorder, measurement invariance, phonological awareness, rapid automatized naming, phonological loop

Educational Impact and Implications Statement

When it comes to diagnosing reading disorder, an assessment of phonological processing is often part of the diagnostic procedure in order to better understand the child's strengths and weaknesses in the phonological domain. With respect to the validity of such diagnostic assessments, it is a crucial aspect of test fairness to know whether or not the administered measurement instruments work equally well across the subgroups for which they are used. Using so-called invariance testing, this study showed that common measures of phonological processing have the same underlying meaning for children with RD and typical learners. This finding is important, because it implies that the results obtained from such a diagnostic assessment can be interpreted in the same way irrespective of the children's group membership.

Phonological Processing in Children with Specific Reading Disorder versus Typical Learners:

Factor Structure and Measurement Invariance in a Transparent Orthography

It is well established that *phonological processing*—the ability to utilize the sound structure of oral language while processing written language (Wagner & Torgesen, 1987)—plays an essential role in the acquisition of reading and spelling skills. For instance, phonological processing contributes significantly to emergent literacy skills even when controlling for other crucial factors such as intelligence level (e.g., Babayiğit & Stainthorp, 2011; Schneider & Näslund, 1993), vocabulary and letter knowledge (e.g., Babayiğit & Stainthorp, 2011; Furnes & Samuelsson, 2011) or grammar skills (e.g., Nikolopoulos, Goulandris, Hulme, & Snowling, 2006). Moreover, longitudinal studies (e.g., Boscardin, Muthén, Francis, & Baker, 2008; Lambrecht Smith, Scott, Roberts, & Locke, 2008; Schneider & Näslund, 1993; Torppa et al., 2013; Wagner, Torgesen, & Rashotte, 1994) have demonstrated that phonological deficits are highly persistent over the childhood years and are causally related to the development of a reading disorder. In fact, it is by now widely accepted that a *phonological core deficit* underlies the cognitive manifestation of a reading disorder (Stanovich, 1988).

Compared to the vast amount of correlational and longitudinal studies conducted in the field, few have been dedicated to the structure of phonological processing. Quite recently, however, an increased interest can be found among researchers in uncovering the nature and dimensionality of phonological processing and examining the extent to which its conceptualizations are transferable to different (sub-)populations. So far, this branch of research has mainly focused on the question to what extent the latent structure of phonological processing is transferable to different orthographies (e.g., Dutch: de Jong & van der Leij, 1999; Latvian: Sprugevica & Høien, 2004; Spanish: Anthony et al., 2006) and as to whether it applies to younger and older children alike (e.g., Anthony, Williams, McDonald, & Francis, 2007). Interestingly, despite the crucial role phonological processing plays in current models of reading disorder no study has yet examined closely whether or not affected children show the same structure of phonological processing as their typically achieving peers. Therefore, the objective of this study was to determine and compare the factor structure and measurement invariance of phonological processing skills across these two groups of children.

On the Structure of Phonological Processing

Wagner and Torgesen (1987) developed an influential model of phonological processing which distinguishes three components: *Phonological awareness* (PA) describes a person's sensitivity to speech sounds and comprises the ability to analyze and manipulate the phonetic structure of spoken language. In this way, PA helps the beginning reader and speller to establish the crucial mapping between letters and sounds, necessary in acquiring an alphabetic script. *Phonetic recoding in working memory* (also referred to as the *phonological loop*, PL) describes the temporary storage of verbal and acoustical information in working memory. For instance, during sentence reading the PL retains an acoustical representation of the words and thereby helps the reader to derive content and meaning. Likewise, the spelling process requires an acoustic representation of the to-be-written word in the PL. Finally, *phonological recoding in lexical access* (preferably assessed by a task called *rapid automatized naming*, RAN) refers to how efficiently phonological information can be retrieved from the mental lexicon of long-term memory. More precisely, lexical access describes the mechanism by which a written word or another visual input leads to the rapid activation of its lexical entry through the process of phonological recoding (Wagner, 1986). Obviously, children with a large sight vocabulary who rapidly retrieve entire words are able to read and spell with greater efficiency than children who use an effortful letter-by-letter decoding strategy. Conceptualized in this way, RAN is considered to be a measure of the efficiency of visual-to-verbal recoding in the mental lexicon (cf. Wagner, 1986). Relatedly, Moll and colleagues (2009) examined the RAN–literacy relationship in German and found evidence to suggest that RAN in this language may best be conceptualized as the automaticity of visual-verbal integration. Nevertheless, any conceptualization of RAN has to acknowledge that the naming of visual items involves a broad array of underlying processes: In addition to phonology, RAN also requires attention and inhibition as well as orthographic processing and general processing speed (cf. Kirby, Georgiou, Martinussen, & Parrila, 2010). Thus, although this study conceptualized RAN mainly as the phonological subcomponent that is responsible for visual-verbal integration, we additionally consider other theoretical explanations of the RAN–literacy relationship.

Wagner, Torgesen and colleagues (1993, 1994) further specified their model in later publications. For instance, they suggested that PA may actually consist of two discrete yet related factors, namely the ability to blend together phonemes or words (*phonological synthesis*) and the ability to identify and manipulate particular phonemes within words (*phonological analysis*). Although certainly this separation is of conceptual value, subsequent studies have supported neither this nor other multi-dimensional theories of PA (e.g., Anthony & Lonigan, 2004; Papadopoulos, Schatschneider, Francis, Foorman, Fletcher, & Mehta, 1999; Spanoudis & Kendeou, 2009; Vloedgraven & Verhoeven, 2009). As a result, PA is nowadays generally considered to be a unitary ability. Moreover, Wagner et al. (1993, 1994) proposed to further subdivide lexical access into processes related to isolated versus serial naming. However, isolated naming has not been proven a unique predictor of literacy skills once serial naming is controlled (e.g., Logan, Schatschneider, & Wagner, 2011) and has thus been of only minor importance in recent studies of reading. In line with these findings, the present study built on the originally proposed model that consists of three rather than five phonological processing components.

There is evidence that those phonological processing skills are related to each other in different ways: Statistical associations between PA and the PL are generally in the medium-to-high range, whereas their relations with RAN are often considerably smaller or even nonsignificant (see Norton & Wolf, 2012, for a review).

Given that Wagner and Torgesen's (1987) model was developed for English (an opaque orthography), one objective of the present study was to investigate whether the proposed structure also applies to the transparent orthography of German. In view of the differences between transparent and opaque orthographies, the question arose to what extent the findings for English can be generalized to other orthographies (e.g., Aro & Wimmer, 2003; Smythe et al., 2008; Ziegler, Perry, Ma-Wyatt, Ladner, & Schulte-Körne, 2003). By now, there is in fact some evidence from cross-language studies suggesting that the manifestation of reading disorder is not universal but depends on the special characteristics of the underlying orthography (Landerl, Wimmer, & Frith, 1997; Vellutino et al., 2004).

Specific Reading Disorder

Specific reading disorder (hereinafter just referred to as reading disorder; F81.0) is a developmental learning disorder listed in the *International Classification of Diseases* (*ICD-10)* of the World Health Organization (WHO, 2011). The main feature according to definition is a significant and unexpected impairment in the development of reading and spelling skills: The learning problems are *significant* in that the child's performance is substantially below the level expected for the child's grade level; and they are *unexpected* because they contradict the child's intellectual potential. This uncoupling between intelligence and academic achievement (referred to as the *IQ-discrepancy*) is at the heart of the medical definition (e.g., Ferrer, Shaywitz, Holahan, Marchione, & Shaywitz, 2010). Specifically, the IQ-discrepancy has fueled the notion that reading disorder is of neurobiological origin and arises from distinct and specific cognitive dysfunctions which are presumably but not exclusively located in the phonological domain (e.g., Lyon, Shaywitz, & Shaywitz, 2003). In other words, the medical definition of learning disorder argues that children who fulfill the IQ-achievement discrepancy criterion are qualitatively distinct from normal readers on the one hand, as well as from poor readers on the other hand, whose reading problems are in line with IQ expectations (e.g., Meyer, 2000).

Regarding the symptoms and manifestations of this disorder, ICD-10 acknowledges some differences between orthographies. While in English the problems tend to center on reading accuracy, there is by now ample evidence that in transparent languages such as German the main symptoms concentrate on slow and dysfluent reading (e.g., Landerl, Wimmer & Frith, 1997; Wimmer & Schurz, 2010). Moreover, since word recognition serves as a bottleneck for higher-order reading skills, the children may also exhibit additional problems in text comprehension (e.g., Peterson & Pennington, 2015).

Phonological Processing in Children with Reading Disorder

Several meta-analytic studies (Kudo, Lussier, & Swanson, 2015; Melby-Lervåg, Lyster, & Hulme, 2012; Swanson, 2012; Swanson & Hsieh, 2009) and literature reviews (Mody, 2003; Vellutino, Fletcher, Snowling, & Scanlon, 2004) come to the conclusion that children and adults with reading disorder exhibit difficulties with all three components of phonological processing. Traditionally, most of the pertinent studies have been concerned with observed rather than latent variables and differences between groups are therefore mainly interpreted as functional deficits (cf. Schuchardt, Roick, Mähler, & Hasselhorn, 2008). That is, when children with reading disorder perform lower on phonological tasks than their typically achieving peers, it is generally assumed that those performance differences are due to quantitative deficits in phonological processing. Yet, poorer performance on manifest measures may just as well result from structural differences in phonological processing: If the separation of phonological processing into three highly specialized components were not evident in children with reading disorder, then PA, PL and RAN would not contribute as much unique variance to the children's literacy development as they do in typical learners. From a theoretical point of view, this assumption of structural differences logically results from the medical definition, which conceptualizes reading disorder as a distinct category rather than the lower end of the ability continuum.

However this view has been challenged, raising the claim for a dimensional re-conceptualization of reading disorder (e.g., Branum-Martin, Fletcher, & Stuebing, 2012; Francis et al., 2005). Historically, Linda Siegel and Keith Stanovich were among the first who extensively criticized the IQ-discrepancy criterion: Comparing IQ-discrepant with non IQ-discrepant poor readers on a range of reading and cognitive measures, they found only little support for differences between the two groups, thereby questioning the usefulness of the medical definition (e.g., Siegel, 1989, 1994; Stanovich, 1994a, 1994b). Since then, there is an ongoing debate as to whether the cognitive differences between children with reading disorder and typical learners are quantitative or qualitative in nature (e.g., Coghill & Sonuga-Barke, 2012; Compton, Fuchs, Fuchs, Lambert, & Hamlett, 2012). While the former view suggests that the cognitive characteristics of children with reading disorder differ from normal reading only in level and degree, the traditional medical view assumes that the cognitive profiles also differ in pattern and kind.

To address this issue, a construct validation study that examines the equivalence of the underlying cognitive structure across the two groups seems reasonable, as this may provide empirical evidence in favor of either the dimensional or categorical conceptualization of reading disorder. Moreover, given its crucial role for the development of reading and spelling, phonological processing is a cognitive source for which structural differences can be plausibly expected. Interestingly, to the best of our knowledge the assumption of structural equivalence of phonological processing has not yet been tested empirically. This constitutes a lack in current research because Schuchardt et al. (2008) and others (e.g., van de Schoot, Lugtig, & Hox, 2012) emphasized that results in a measurement instrument can only be validly compared across qualitatively distinct entities if the underlying latent constructs and the test properties do not differ systematically for these groups. If, however, group membership moderates either the nature of phonological processing *or* the relationship that exists between the observed test scores and the latent constructs, valid interpretation of group differences is more difficult.

Conversely, if the conceptualization of reading disorder was quantitative in nature and just captured the lower end of the ability continuum, current practice would not be so problematic as–in this case–measurement invariance probably holds. Thus, instead of assuming that the structural conceptualization of phonological processing applies to children with reading disorder in the same way as for typical learners, we argue that it is important to test for measurement and structural invariance among these two groups of children.

In fact, there are several circumstances that may suggest a difference in the latent structure of phonological processing in children with reading disorder and typical learners:

1. *Phonological processing may not follow a tripartite structure in children with reading disorder*

According to Carroll’s (1993) hypothesis of proficiency-based divergence in latent factors, children might differ not only quantitatively in their actual cognitive performance levels, but also qualitatively in how their cognitive abilities are structured and specialized. Particularly, this hypothesis supposes that in young or low-proficient children cognitive abilities tend to be minimally differentiated and might therefore be captured by only one or few latent factors; however, with increasing age and proficiency the children’s cognitive abilities fan out and become more and more specialized, which in turn produces a more complex factor structure. In line with this hypothesis, there is some evidence that children's phonological processing structure undergoes a developmental change during the first years of formal reading instruction. Specifically, the tripartite structure of phonological processing does not seem to be in place in children below the first and second grade. For instance, in a study conducted by Wagner et al. (1993), only lexical access (assessed by RAN tasks) emerged as a discrete ability in pre-readers, whereas measures of PA and PL were not distinguishable from each other (i.e., they loaded on one and the same factor). Among the second-graders tested in this study, it was however possible to fit the classical phonological processing model with separate factors for PA, PL and RAN. This finding supports the notion that children's phonological abilities become increasingly differentiated when they develop from initial to skilled reading (cf. Lonigan et al., 2009). However, pre-readers who later develop a reading disorder might possibly not undergo this important developmental shift in their phonological processing structure– maybe due to some inherent neurobiological factors. This assumption seems reasonable as reading disorder is considered a developmental disorder of neurobiological origin (e.g., Lyon et al., 2003; WHO, 2011). That is, PA and PL may constitute just one latent factor throughout the children's development and may be thus less differentiated, which may at least partially explain the children's learning problems.

2. *The degree to which the phonological processing components vary may be different for children with reading disorder and typical learners*

The latent constructs of PA and RAN may be differently related to each other in children with reading disorder and typical learners. For instance, in their meta-analysis of 49 correlational studies, Swanson and colleagues (2003) reported a moderate correlation of r = .42 between tasks of PA and RAN among skilled readers. In contrast, among children with poor reading skills the corresponding correlation—although corrected for range restrictions and sample size—was considerably lower ($r = .22$), indicating a lower linear dependence between PA and RAN in this group. These findings correspond with the *double-deficit hypothesis* (Bower & Wolf, 1999). Accordingly, deficits in PA and in RAN constitute two independent sources of reading failure, resulting in different subtypes of struggling readers: Although the majority of poor readers (approx. 50 to 60%) is likely to exhibit deficits in both phonological processing skills (double deficit), others (approx. 25 to 35%) do poorly in either PA or in RAN only (single deficit); few poor readers may not be classified (Bower & Wolf, 1999; Wolf et al., 2002). The existence of those subgroups may lead to lower relationships between latent conceptualizations of PA and RAN and may also lead to a higher variability of phonological processing skills in poor readers as opposed to good readers. Hence, tests of invariance seem especially suitable in order to compare the "true" (error-free) covariations and variance components of phonological processing skills across children with reading disorder and typical learners.

3. *Across the two groups, phonological processing tasks may not function equally well as indicators for the latent constructs*

A particular phonological task may work as a good indicator of phonological processing in one group but not in the other. Statistically speaking, this would be the case if factor loadings varied as a function of group membership. For example, in their study with younger and older preschoolers, Lonigan et al. (2009) as well as Papadopoulos et al. (2012) reported that the extent to which scores on PA tasks were accounted for by the underlying phonological factor differed systematically between the tested age groups. Additional evidence that phonological tasks may not be equally effective across different ability levels comes from Schatschneider et al. (1999). Using item response theory, the authors found that the difficulty and discriminability of various PA tasks was dependent on the children's ability levels. The authors therefore concluded that the diagnostic usefulness of a specific PA task cannot be a priori generalized across the ability continuum. It should thus be born in mind that one and the same phonological processing task may not function equally well across different (sub-)populations. Some subtests may provide only limited information about low-ability children, but much information about high-ability children; whereas other subtests may be more effective in measuring phonological processing among children with low-ability levels. However, to the best of our knowledge no study has yet examined whether this restriction is also evident in children with reading disorder. This is surprising, given that group invariance of factor loadings is an important issue in construct validation: It ensures that the underlying latent construct is measured in the same way across groups and has thus the same underlying meaning (Meredith, 1993; Millsap, 2011).

4. *Measurement error may systematically differ for children with reading disorder and typical learners*

Lastly, the reliability of phonological tasks may systematically differ for both groups of children: Since struggling readers experience greater difficulties with phonological tasks than typical learners, they might get more easily distracted while performing those tasks or they might exhibit greater signs of frustration and motivation loss relative to their typically achieving peers. Likewise, floor effects may also occur more often in children with reading disorder. Obviously, those effects might reduce the reliability of the testing instruments. To rule out the possibility that measurement error systematically differs across groups, it seems therefore necessary to examine whether residual variances of phonological processing tasks (i.e., the observed variance *not* accounted for by the underlying phonological processing factors) are invariant across children with reading disorder and typical learners.

To summarize, rather than assuming that the latent structure of phonological processing applies to children with reading disorder and typical learners alike, we argue that it is important to test for structural and measurement invariance among these two groups of children. Particularly, our study was designed as a construct validation study that aimed at answering the following two questions:

1. *Does phonological processing follow the proposed tripartite structure in children with reading disorder who acquire a transparent orthography?*
2. *If so, is the phonological processing model invariant across children with reading disorder and typical learners?*

Method

Procedure

The study was part of a multi-centric research project that aimed at investigating the interplay between cognitive functioning and school achievement in children with learning disabilities.

Recruitment of participants. All children of the reference group and most children of the reading disorder group (RD group; 92%) were recruited via a screening on school achievement that took place in elementary schools in and around three cities located in the northern and central parts of Germany (viz., Frankfurt am Main, Hildesheim, and Oldenburg). The remaining children with RD were recruited by a counseling center for learning disabilities in Hildesheim. Overall, 3,205 children (age: *M* = 8;7 years, *SD* = 6 months; 49.2% girls) from 134 schools and 280 classes participated in the screening. The children completed standardized school achievement tests of reading, spelling, and mathematics as well as a standardized measure of nonverbal IQ. They were tested in groups at their school over two 90-minute lessons. Given our diagnostic criteria (outlined below), the screening revealed a prevalence rate of reading disorder of about 8.7% (Authors, 2013), which is within the range reported in other prevalence studies (e.g., Dirks, Spyer, van Lieshout, & de Sonneville, 2008; Landerl & Moll, 2010; see also Peterson & Pennington, 2015, for a review). Children fulfilling the criteria for the RD group or the reference group were either invited to take part in the main study or a related study. Of the 527 children invited to the reference group, about 30% decided to participate in one of the two studies. Overall, participating children were mostly comparable to non-participating children with respect to the classification measures and age. Significant but marginal group differences only emerged in reading (*T*-score difference: $\Delta M = 1.50$; $\Delta SD = 0.03$; $\eta_p^2 = .018$) and in mathematics (*T*-score: $\Delta M = 1.61$; $\Delta SD = -0.26$; $\eta_p^2 = .017$); with non-participating children outperforming participating children. For the RD group, return rate was about 34% and again participating children were mostly comparable to non-participating children. Significant but marginal group differences emerged in reading (*T*-score: $\Delta M = 3.05$; $\Delta SD = 0.43$; $\eta_p^2 = .029$) and in spelling (*T*-score: $\Delta M = -1.18$; $\Delta SD = -0.91$; $\eta_p^2 = .011$), with non-participating children performing slightly better in reading but poorer in spelling.

Assessment of cognitive functioning (main study). Measures of phonological processing were only administered to students participating in the main study. The assessment took place individually in two sessions, each lasting up to ninety minutes. Student research assistants tested the children in schools or in the universities' laboratories.

Parental informed written consent was obtained for all children prior to testing. Participation was voluntary and consent could be withdrawn at any time without giving reasons.

Participants

The sample of the main study included 209 third graders. Of these, 100 children belonged to the reference group and 109 to the RD group. Classification was based on norm-referenced measures (standard scores). The criteria for the reference group were as follows: Children's achievement scores in reading, spelling and mathematics were at grade expected levels, with standardized scores of at least $T \geq 45$ (*T*-score: $M = 50$; $SD = 10$). In contrast, children of the RD group showed achievement scores in reading and/or spelling that were below grade expected levels (i.e., at least 1.0 *SD* below the normed reference group's mean; T-score ≤ 40), whereas their performance in mathematics was grade appropriate ($T \geq 40$ and at least 5 T-points above the child's reading and spelling scores). In line with ICD-10 (WHO, 2011), we additionally applied an IQ-achievement discrepancy criterion to the definition of reading disorder: In particular, the children showed a critical discrepancy of at least 1.2 *SD*s between their nonverbal IQ and their literacy achievement. Also, all children showed at least average intelligence (IQ $\geq$ 85).

Since the cut-off criteria used in the classification of reading disorder are rather heterogeneous (e.g., Elliott & Grigorenko, 2015), we want to outline the rationale for the criteria used in the present study: In general, a norm-referenced cut-off score of $T < 40$ for the low achievement criterion and of 1.2 *SD*s for the IQ-discrepancy criterion correspond to the recommended diagnostic guidelines (Strehlow & Haffner, 2002) and they are most frequently used in German educational and clinical settings (Hasselhorn, Mähler, & Grube, 2008; Klicpera, Schabmann, & Gasteiger-Klicpera, 2010). That is, by applying those cut-off scores, our sample best represented the subpopulation of school children in Germany commonly referred to as having a reading disorder.

Table 1 shows the descriptive characteristics of the sample as a function of group. A multivariate analysis of variance (MANOVA) was performed to check whether the groups differed with respect to the classification measures. For these and the following statistical analyses, the alpha level was set at $p = .05$. Effect sizes are reported using partial eta-squared (η_p^2) classified by Cohen (1988) as small (.01–.05), medium (.06–.13), and large ($\geq$ .14) effects. The multivariate main effect was statistically significant, Wilks's $\Lambda = .29$, $F(4, 204) = 125.33$, $p < .001$, $\eta_p^2 = .71$; and was therefore followed up with separate univariate ANOVAs: No statistically significant differences between groups were found for nonverbal intelligence, $F(1, 207) < 1$, $MSE = 108.69$, $p = .402$, $\eta_p^2 < .01$; and for mathematical skills,

$F(1, 207) < 1$, $MSE = 36.65$, $p = .493$, $\eta_p^2 < .01$. However, as could be expected due to the sampling procedure, groups differed significantly in reading skills, $F(1, 207) = 193.54$, $MSE = 50.34$, $p < .001$, $\eta_p^2 = .48$; and in spelling skills, $F(1, 207) = 263.07$, $MSE = 36.61$, $p < .001$, $\eta_p^2 = .56$. There were no statistically significant group differences with respect to chronological age as indicated by an ANOVA, $F(1, 207) = 2.59$, $MSE = 28.84$, $p =.109$, $\eta_p^2 = .01$. Whereas sex distribution was balanced in the reference group (50 boys; 50 girls), there were more boys than girls in the RD group (77 boys; 32 girls), $\chi^2 (1, N = 209) = 9.32$, $p = .002$. This is in line with prevalence studies showing that learning disorders in the literacy domain are generally more frequent in boys than in girls (e.g., Moll, Bruder, Kunze, Neuhoff, & Schulte-Körne, 2014). Lastly, the RD group and the reference group were balanced with respect to the children's home towns, $\chi^2 (2, N = 209) < 1$, $p = .952$, $\eta^2 = .22$ (reference group: 46% Frankfurt, 29% Hildesheim, 25% Oldenburg; RD group: 44% Frankfurt, 29% Hildesheim, 27% Oldenburg).

[Please insert Table 1 about here]

Reading and Writing Curriculum in German Elementary Schools

Due to the transparency of the German orthography, reading and spelling instruction often follows a synthetic phonics-based teaching approach. That is, children are systematically taught all existing grapheme-phoneme relations and learn how to derive word pronunciation on the basis of these conversion rules. Likewise, there is a high prominence on phoneme-grapheme relations in writing instruction, at least in the first years of schooling: Children are usually taught to segment the sound sequences of words into individual phonemes and are encouraged to write down the corresponding graphemes (e.g., Niedersächsiches Kultusministerium, 2006).

Tasks and materials

Classification measures. To obtain an estimate of general cognitive ability, children completed the German version of the *Culture Fair Intelligence Test 1* (CFT 1; Cattell, Weiß, & Osterland, 1997). The CFT 1 is a nonverbal measure used as an indicator of fluid intelligence; it comprises five subtests that can be classified according to two parts. The first part consists of two measures, which assess perceptual speed, visual attention and visuomotor ability. The second part consists of three subtests, which examine deductive reasoning skills and figural thinking. All items are dichotomously scored and then combined to an overall performance score. Split–half reliability is .90 and .92 for the age groups studied. The validity of the CFT 1 has been well established and the manual states reasonable levels of convergent and criterion validity as well as factorial validity.

We tested the children's reading skills with the *ELFE 1–6* (Lenhard & Schneider, 2006), which consists of three subtests. The first subtest assesses decoding speed by means of a picture-word-matching procedure: Each of the 72 items consists of one picture and four written words. The children's task is to identify the word that corresponds to the picture. The children have three minutes to work on as many items as possible. The second subtest measures reading at sentence level and consists of 28 unrelated sentences, each missing a word. Out of a set of five words, the children are asked to identify the word that completes each sentence correctly. Children have three minutes to work on as many items as possible. The third subtest assesses reading at text level and requires the children to answer multiple-choice questions in response to short narratives. Children have seven minutes to complete the task. The ELFE 1–6 is designed as a speed test rather than a power test so that item difficulty (especially for subtest 1 and 2) is generally low. Consequently, children rarely make mistakes and skill differentiation is mainly based on reading speed. A reading test covering reading speed and to a smaller extent reading comprehension was used rather than a reading accuracy test, as reading accuracy is usually high in transparent orthographies (e.g., Landerl, 2001;

Landerl et al., 1997), and it consequently does not distinguish sufficiently between good and poor readers. Items of the three subtests are dichotomously scored and then combined to an overall performance score. Internal consistency of the ELFE 1–6 is high with values between .92 and .97. The manual reports adequate levels of convergent validity (e.g., correlations of r = .48 to .79 with other standardized reading tests) and satisfactory levels of criterion validity (e.g., correlations of about r = -.76 with the children's grades in German).

To assess spelling achievement, the *WRT 2+* (Birkel, 2007), a German spelling test for second and third graders was administered. This test requires children to spell 43 dictated words embedded in short sentences. Items of the WRT 2+ are dichotomously scored; the dependent variable is the number of correct spellings. Internal consistency for this measure is reported as high with *Cronbach's* α = .94. The manual reports satisfactory levels of convergent validity (e.g., correlations of r = .65 to .85 with other standardized spelling tests) and adequate levels of criterion validity (e.g., correlations of r = -.65 to -.69 with the children's grades in German).

To control for co-occurring learning disabilities in mathematics, children completed the *DEMAT 2+* (Krajewski, Liehm, & Schneider, 2004), a curricular-valid test of basic arithmetic, magnitude, and geometry. Items of this test are dichotomously scored; the dependent variable is the number of problems solved correctly. Internal consistency of the DEMAT 2+ is reported as .91 for third graders. The DEMAT 2+ shows moderate to strong levels of prognostic validity (r = .63 to .67), convergent validity (r = .53 to .67) and criterion validity (r = .66).

Measures of phonological processing. Because this study can be construed as a construct validation study, we particularly used those types of tasks that are commonly considered to be typical measures of phonological processing. Following the suggestions of Anthony et al. (2007) as well as Lonigan et al. (2009), we further deemed it necessary to only use those measures that are not related with two or more phonological processing skills at the same time to prevent confounding of constructs. For instance, a nonword repetition task was therefore not applied in the present study. Although nonword repetition was originally introduced as a relatively pure measure of the PL (Gathercole & Baddeley, 1993), researchers have since argued that the mechanisms underlying repetition of nonword tasks are by far more complex (e.g., Archibald & Gathercole, 2007; Gathercole, 2006) and less well understood (cf. Metsala, 1999) than in serial recall measures such as word span or digit span. Furthermore, it is by now widely accepted that (other than serial recall) nonword repetition (a) assesses the phonological *sound quality* of item storage rather than the overall *storage capacity* of the PL (e.g., Hasselhorn, Grube, & Mähler, 2000) and (b) it is thus in a conceptual sense highly related to PA (see Bowey, 1997; Gathercole, 2006, for reviews). Specifically, storing and repeating a nonword for which no lexical entry exists requires children to identify and segment the presented phoneme structure in a deep manner, because only in this way the item can be maintained correctly in the PL. In fact, empirical studies (e.g., Metsala, 1999) have demonstrated that nonword repetition is not only accounted for by phonological storage capacity, but also by phonological sensitivity as assessed in PA.

Along with the description of subtests, we provide sample-based reliability estimates. Overall, the internal consistency of the measures was around .70 to .90, which is within an acceptable and common range for basic research (cf. Nunnally & Bernstein, 1994).

Phonological awareness (PA). The *Test of Basic Competencies for Reading and Spelling* (BAKO 1–4; Stock, Marx, & Schneider, 2003) was used to assess the children's PA on phoneme level. Of the seven subtests of the BAKO 1–4, we selected those three subtests that showed the best item characteristics, namely Vowel Substitution, Vowel Length, and Phoneme Reversal. That is, not used in the present study were the BAKO 1–4 subtests Nonword Segmentation, Phoneme Inversion, Sound Categorization, and Phoneme Deletion.

The manual reports expected and satisfactory levels of criterion validity by correlating the BAKO 1–4 with a range of widely used achievement tests (e.g., moderate to high correlations with measures of reading and spelling; non-significant to low correlations with nonverbal IQ). Items of the BAKO 1–4 are dichotomously scored.

In the Vowel Substitution test, the child's task is to substitute all /a/ vowels in a given word by an /i/ vowel (e.g., Sand → Sind). This test is based on vowel phonemes rather than consonant phonemes, because in a lot of languages (including German) vowel changes are used to indicate the tenses of irregular verbs and are thus a crucial phonological marker (cf. Wimmer, Landerl, Linorter, & Hummer, 1991). The test consists of 12 items (8 words, 4 pseudowords), which range between two and four syllables in length. Three practice trials are presented prior to testing. Internal consistency of this measure (based on the *Kuder–Richardson* formula 20 due to the dichotomy of the item scoring) was $KR_{20} = .78$ for the RD group and $KR_{20} = .76$ for the reference group.

The Vowel Length subtest which has mainly been used in transparent orthographies is a modification of Bradley and Byrant's (1985) Sound Categorization/Oddity Detection task and assesses vowel *duration* instead of vowel *identity*. Accordingly, this test is phonologically more demanding than its original and it thus works well with third graders (Stock, Marx, & Schneider, 2003). In contrast, discrimination of vowel *identity* (as assessed in the conventional Sound Categorization task) is mastered early by children acquiring transparent orthographies such as German (Landerl, 2003). Moreover, a measure of Vowel Length is of interest because being able to correctly perceive and discriminate vowel lengths in spoken German is an important phonological skill, which is, for example, required in learning the difficult German spelling rules to mark short and long vowels (cf. Landerl, 2003). The child is presented four pseudowords, all of which contain the same vowel phoneme in the middle. Yet, one of the pseudowords differs from the others in vowel length (e.g., /re:m/ - /fe:r/ - /nɛl/ - /be:f/). The child's task is to identify the item that does not match the others. The subtest consists of two practice trials and ten test trials. Internal consistency of this measure was KR_{20} =.69 for the RD group and KR_{20} =.79 for the reference group.

In the Phoneme Reversal subtest, the child's task is to pronounce a given (pseudo)word in reversed order (e.g., ruf → fur). Of the 18 test items, ten consist of pseudowords, four of which turn to real words after reversal. Again, two practice trials are presented prior to testing. Internal consistency of this measure was KR_{20} = .88 for the RD group and KR_{20} = .89 for the reference group.

Items of the BAKO 1–4 were presented audibly via computer and the examiner recorded the child's verbal responses on a protocol sheet. Subtest presentation was stopped once the child answered three subsequent items incorrectly. The dependent variable was the number of correct trials up to the point at which the subtest was stopped.

Phonological loop (PL). Three subtests of the computerized and adaptive *Working Memory Test Battery for Children aged Five to Twelve Years* (AGTB 5–12; Hasselhorn et al., 2012) were administered to assess phonetic recoding in short-term memory. The AGTB 5–12 is a German computerized test battery, which assesses working memory skills according to Baddeley's (1986) multicomponent model.Construct validity of the AGTB 5-12 was established in a large study with 1,669 children (Michalczyk, Malstädt, Worgt, Könen, & Hasselhorn, 2013). Furthermore, the AGTB 5–12 demonstrates significant criterion validity with respect to reading and spelling tests (Hasselhorn et al., 2012).

In the Digit Span task, increasing sequences of different digits are presented audibly at the rate of one digit every 1.5 s. No digit appears twice in a particular sequence. The child's task is to repeat the sequence orally in the same serial order as presented. The sample Cronbach's alpha coefficient was .90 for the RD group and .92 for the reference group.

Similarly, the Word Span task requires the serial repetition of high-frequency words, which are presented audibly at the rate of one word every 1.5 s. Word sequences are constructed out of nine phonologically and semantically dissimilar German nouns. Each word appears only once within a particular sequence. There are two versions of the task—one using monosyllabic and one using trisyllabic words—resulting in separate span scores for short and long words, respectively. For the monosyllabic word span, the sample Cronbach's alpha coefficient was .89 for the RD group and .92 for the reference group. For the trisyllabic word span, Cronbach's alpha was .79 and .89, respectively.

All three tasks are span measures with an adaptive testing procedure. They consist of ten trials, divided into five testing blocks with two trials each. The first testing block starts with a three-item sequence, and sequence length is adjusted after each response: If the child recalls the presented trial correctly, the sequence length of the subsequent trial increases by one item. If, however, the child's recall is incorrect, the sequence length of the next trial decreases by one item. In the remaining four testing blocks, sequence length is adjusted more conservatively as follows: If the child recalls both trials of the testing block correctly, the span length of the next block increases by one item. If, however, the child recalls both trials incorrectly, the span length decreases by one item. If recall is incorrect for only one of the two trials, the span length remains the same. The calculation of the span score is based on the mean performance in the last four testing blocks: For each correct response, the child receives a score that corresponds to the span length. For instance, if the child correctly recalls a five-item sequence, he or she receives five points. A false response is assigned the span length decreased by one item (e.g., incorrect repetition of a five-item sequence results in four points only).

Rapid automatized naming (RAN). This task measured lexical access from long-term memory. The RAN task consisted of two alphanumeric subtests, which assessed naming speed for digits and letters, with items adapted from the *3 DM Dyslexie* (Blomert & Vaessen, 2008). In addition, two non-alphanumeric subtests assessed naming speed for colors (red, yellow, blue, green, and black; items adapted from Denckla & Rudel, 1976), and objects (car, fish, hammer, dog, candle; items adapted from Blomert & Vaessen, 2008). In each subtest, items are arranged randomly in five rows of 10 on a white paper (size: 41.0 x 29.5 cm). The child's task is to name all items as quickly as possible while making as few errors as possible. Naming time (in seconds) served as the dependent variable, that is, lower scores indicated higher performance. Each subtest was preceded by a short practice trial (i.e., two rows à five items) in order to familiarize the child with the material. The sample Cronbach's alpha coefficient was .73 for the RD group and .74 for the reference group.

Statistical Analyses

First research question: Factor structure of phonological processing. To investigate the factor structure of phonological processing in children with RD versus typical learners, two alternative models were tested:

(a) a two-factor oblique model in which the RAN items were captured by one of the factors, whereas the PA and PL items were captured by the other factor, and
(b) a three-factor oblique model with separate latent factors for PA, PL, and RAN.

The two-factor model is nested under the three-factor model, because it can be obtained by restricting the intercorrelations between the latent factors of PA and PL to 1. Models were tested in Mplus 7.11 (Muthén & Muthén, 1998-2012) using maximum likelihood estimation with robust standard errors (MLR). The MLR estimator treats missing values in a full information approach and can be applied to non-normal distributed data (Wang & Wang, 2012). Following Hu and Bentler's (1999) criteria of model fit, a good fit to the data was indicated by (a) a comparative fit index (CFI) with values of at least .95, (b) a root mean square error of approximation (RMSEA) with values of .06 or less, and (c) a nonsignificant χ^2–test. In addition, we report two information criteria indices, namely the Akaie's information criterion (AIC) and the Bayesian information criterion (BIC), where smaller values represent a better fitting model.

Second research question: Measurement invariance of phonological processing. In construct validation studies, the following invariance tests are generally considered necessary: invariance of factor loadings, variances, and covariances (e.g., Byrne, 2012; Vandenberg & Lance, 2000). Thus, the next set of analyses determined whether the measurement parameters (i.e., the factor loadings and residual variances) as well as the structural parameters (i.e., the factor variances and covariances) of the established factor model operated equivalently across both groups of children. Hence, a sequence of increasingly restrictive multi-group models was tested, with equality constraints imposed to the covariance structure in a hierarchical manner. In each testing step, we used nested model comparison to assess whether the more restrictive model was preferable to the less restrictive but slightly better fitting comparison model. Specifically, nested model comparison was twofold: (1) Since we used the MLR estimator, changes in model fit were examined with the Satorra-Bentler scaled χ^2–difference test (Δ SB-χ^2; Satorra & Bentler, 2001) rather than the likelihood ratio (LR) test. A nonsignificant value of the Δ SB-χ^2–statistic implies that the restrictive model fits the data just as well as the less restrictive comparison model (e.g., Wang & Wang, 2012). Thus, the hypothesis of parameter invariance can be accepted. (2) The CFI difference value (Δ CFI) served as a second indicator for multi-group invariance: A difference value less than or equal to 0.01 between the restricted and the comparison model suggests parameter invariance (Cheung & Rensvold, 2002).

Results

Data Screening

Table 1 shows descriptive statistics of the phonological processing measures as a function of group. Prior to the main analyses, we evaluated whether the data met basic assumptions of confirmatory factor analyses (CFA). In particular, none of the zero-order correlations between the manifest variables was above .80 (Table 2), indicating no problem with multicollinearity. In addition, the data were checked for univariate outliers that we classified in terms of cases more than 3.5 *SD*s from the sample's means: Of the 2067 values in the dataset, seven values were univariate outliers (four children of the RD group; three children of the reference group). These values were deleted from further analyses. No cases were identified as multivariate outliers with $p < .001$ through Mahalanobis distance. There was no evidence that the assumption of univariate normality distribution was violated, since all measures showed skewness less than 3 and kurtosis less than 4. Yet, Mardia's test of multivariate normality (Mardia, 1974) revealed a violation of the multivariate skewness assumption, whereas assumption of multivariate kurtosis was met for both groups. Model estimation was therefore based on the MLR estimator, which offers χ^2-test statistics and standard errors that are robust to non-normal data (Wang & Wang, 2012).

[*Please insert Table 2 about here*]

First Research Question: Measurement Model of Phonological Processing

For both groups, the two-factor oblique model provided only poor fit to the data, as can be seen in Table 3. We then tested the three-factor oblique model: Whereas this model provided an excellent fit to the data of the RD group, results of the reference group revealed a poor fitting model. Note, for example, that the χ^2 –value was highly significant in the reference group. In relatively small samples such as ours a significant χ^2 –value is always of concern (Kline, 2016). We therefore consulted modification indices (MI), which yielded evidence that the residuals between the object naming task and the color naming task covaried highly in the reference group (MI > 22). This covariation may be due to method effects, as both measures are based on non-alphanumerical stimuli. Further, the object naming task, which was administered first, might have primed the color naming: Although the object naming task consisted of black and white drawings, some children might have associated prototypical colors with the objects. As Byrne (2010) demonstrated, inclusion of correlated residuals (even if included in one of the groups only) is of no concern in testing for multi-group invariance. We thus included this additional path in the model in order to account for potential method effects or other sources of systematic variance across tasks[1]. By this means, the model was significantly improved as indicated by a chi-square difference test, Δ SB-χ^2 (1, N = 100) = 21.26, p < .001 and the overall fit of this respecified model was excellent (Table 3).

[Please insert Table 3 about here]

For both groups, the three-factor solution is shown in Figure 1: The PA factor captures variance that relates to the children's awareness of phonological sound patterns and their ability to discriminate or manipulate the phonemes in spoken language. The PL factor captures the children's ability to retain acoustical information in verbal working memory. Finally, the RAN factor captures the children's ability to rapidly retrieve phonological information from long-term memory. Having established the three-factor phonological processing model as the best fitting model for each group, this model served as a baseline for the subsequent multi-group CFAs. Unstandardized factor loadings and item uniquenesses are presented in Table 4.

[Please insert Figure 1 about here]

[Please insert Table 4 about here]

Second Research Question: Factorial and Measurement Invariance of Phonological Processing

Table 5 summarizes the results concerning group invariance. First, we tested for configural invariance and thus determined whether the number of factors as well as the pattern of free and fixed factor loadings was equal across groups. More precisely, the three-factor model that was established as a baseline model for both groups *separately* was now tested for the two groups *simultaneously*. Since, at this stage, no equality constraints are imposed, the configural model constitutes the least restrictive multi-group model. As shown in Table 5, the configural model yielded an excellent fit to the data. Thus, we can conclude that the factor pattern of phonological processing is invariant across both groups.

Second, we examined whether the linear relationships between the 10 indicators and the three underlying factors were invariant across groups (i.e., metric invariance). We therefore included equality constraints on the factor loadings in the multi-group model. Although these constraints led to a slight decrease in overall model fit, the model still provided an excellent fit to the data. Moreover, neither the χ^2–difference test (Δ SB-χ^2 = 11.19, Δ *df* = 7, *p* = .130), nor the CFI difference value (.01) suggested rejecting this model in favor of less parsimonious configural model. Thus, invariance of factor loadings was confirmed, which indicates that the phonological processing factors are measured in the same way in both groups. In other words, the extent to which the phonological processing factors accounted for performance differences in the observed variables did not differ for children with RD versus typical learners: For both groups, a one unit change in the underlying factor scores led to the same degree of change in the observed variables.

Third, to determine whether error variances of the phonological processing model were equal across groups, equality restrictions were additionally imposed on the residual variances of the observed indicators. Again, there was a slight decrease in overall model fit. However, neither the χ^2–difference test (Δ SB-χ^2 = 14.99, Δ *df* = 10, *p* = .133), nor the CFI difference value (.01) suggested rejecting this model in favor of the less parsimonious metric model. Thus, invariance of error variances was confirmed, which suggests that for each indicator the amount of variance that is explained by the underlying factors did not differ between children with RD and typical learners. In other words, phonological processing was measured with the same amount of error in both groups.

Fourth, we examined whether the distribution of the underlying factor scores differed between groups. Therefore, equality constraints were additionally imposed on the factor variances. Again, neither the χ^2–difference test (Δ SB-χ^2 = 3.99, Δ *df* = 3, *p* = .263), nor the CFI difference value (< .01) suggested rejecting this model. Thus, we can assume children with RD and typical learners showed similar performance variation in their underlying phonological processing skills.

Last, to determine whether the relationships among the phonological processing factors were invariant across groups, we additionally imposed equality constraints on the factor covariances. Again, nested model comparison revealed a nonsignificant χ^2–difference test (Δ SB-χ^2 = 2.13, Δ *df* = 3, *p* = .546), and a CFI difference value less than .01. This result suggests that the degree to which the phonological processing factors covaried with each other did not differ between groups.

[Please insert Table 5 about here]

Discussion

Although children with specific reading disorder have often been compared to typically achieving learners on various phonological processing tasks, to our knowledge no study so far has examined whether the structure of phonological processing applies to both groups of children alike. Yet, this question is important, because the medical definition of reading disorder implies that this disorder is qualitatively distinct from normal reading (e.g., Meyer, 2000). Measurement and structural invariance is then a necessary precondition in order to validly compare the two groups and ensure measurement of the same construct in both groups. Therefore, this study examined the invariance of phonological processing between children with reading disorder and their typically achieving peers. To this end, 109 third graders with RD and 100 third graders without any learning problems completed a comprehensive test battery assessing PA, RAN, and PL; we then conducted invariance tests to determine and compare the structure of these skills across both groups. The study yielded three central findings:

First, in both ability groups the model that fitted the data best was a three-factor oblique model with separate factors for PA, PL, and RAN. That is, children's phonological processing involves distinct abilities for the phonetic analysis, the short-term storage and the long-term retrieval of oral language. Our finding that a tripartite structure underlies phonological processing in normally achieving school children replicates previous studies in the field (e.g., Sprugevica & Høien, 2004), but it also adds to existing research by demonstrating that Wagner and Torgesen's (1987) phonological processing model also applies to children whose language is German, an orthography with transparent grapheme-to-phoneme correspondences. This is further evidence for the notion that the general nature of phonological processing is relatively universal, although differences across orthographies may exist in the *relativ*e contribution of those skills to emergent literacy (e.g., Smythe et al., 2008; Vaessen et al., 2010). Further and more importantly, we demonstrated that the tripartite structure of phonological processing also applies to children with reading disorder: The structural organization of their phonological processing skills is as differentiated and developed as in typical learners, albeit at a less efficient level. Overall, this finding is empirical evidence for the common but so far untested assumption that phonological deficits in children with reading disorder reflect functional rather than structural deficits. This finding challenges the medical definition of reading disorder, which implies that children with IQ-discrepant reading problems reflect a qualitatively distinct group (e.g., Meyer, 2000). Rather, our results support a dimensional conceptualization of reading disorder–at least with respect to underlying phonological processing: Deficits in PA, RAN, and PL seem to constitute the lower end on the population continuum rather than being a manifestation of an entirely different or less differentiated phonological processing structure.

Second, we found evidence of measurement invariance. That is, group membership did not moderate the relations between observed test scores and underlying constructs. In particular, we tested the covariance structure of the phonological processing model and found that both the measurement parameters (i.e., the factor loadings and residual variances) as well as the structural parameters (i.e., the factor variances and covariances) were invariant across children with reading disorder and typical learners. The invariance of factor loadings implies that the phonological constructs are measured in the same way in both groups. Further, the invariance of residual variances indicates that phonological processing skills are measured with the same amount of error in both groups. In other words, there is no evidence that aspects influencing reliability such as motivation loss or floor effects are a greater issue in children with reading disorder than in typically achieving children. Invariance of factor variances further suggests that the distribution of underlying phonological processing components did not differ between the two groups. That is, the variance of the phonological factors is neither reduced nor increased in children with reading disorder as opposed to typical learners. Last, invariance of covariances means that the degree to which the phonological processing factors are related to each other did not differ across groups. These results have implications with respect to previous empirical studies: Traditionally, the vast majority of studies on reading disorder comprises small to medium sample sizes due to the enormous efforts that need to be invested in recruiting participants. As a consequence, it is usually not possible to analyze the data with latent models. Group comparisons are therefore usually performed on the manifest level, which, of course, requires strict assumptions about the underlying data. Unlike in latent models, these assumptions usually cannot be checked in manifest analyses. That is, the researchers must simply assume that invariance holds and that group comparisons are thus meaningful. Hence, construct validation studies like the one presented here are of crucial importance as they provide empirical evidence for the assumption of group invariance and thereby underpin the interpretations that are drawn from manifest studies.

Finally, the correlations among the three latent factors are noteworthy. Latent correlations are often referred to as true relationships, because measurement error and task-specific effects are controlled. In both reading groups, the correlations between PA and PL were in the medium range, whereas the respective correlations with RAN were nonsignificant. This correlation pattern corresponds to previous studies (e.g., Norton & Wolf, 2012) and is in line with Wagner and Torgesen’s (1987) theoretical assumption that PA, PL and RAN each represent different facets of phonological information processing. They may therefore make quite specific and unique contributions to developing literacy skills. Consequently, all three phonological processing skills should be targeted in individual diagnostics, because profiles of phonological strength and weaknesses may vary among subgroups of children with reading disorder. Empirically, this idea is supported by studies which found PA and PL to be more related to reading accuracy or spelling, and RAN to be more related to reading speed (e.g., Ennemoser, Marx, Weber, & Schneider, 2012; Moll, Ramus et al., 2014). Moreover, this correlation pattern may be taken as evidence to suggest that PA and PL on the one hand and RAN on the other hand represent independent causes for poor literacy skills. This idea is partly expressed in the double deficit hypothesis (Bower & Wolf, 1999).

In a similar vein, there is some debate in the literature as to whether the low correlations with lexical access stem at least partially from a speed accuracy confound, since

RAN but not PA and PL are measured with speeded items (cf. Wagner et al., 1993). In fact, using a speeded measure for their PA tasks, Vaessen, Gerretsen, and Blomert (2009) demonstrated that only the speeded but not the conventional PA subtests were significantly related to RAN.

Should We Move Toward a Dimensional Conceptualization of Reading Disorder?

Current ICD-10 (WHO, 2011) classification of reading disorder is categorical in nature and expresses the idea that reading disorder is qualitatively distinct from typical reading. Although this distinctiveness hypothesis has been criticized for years (e.g., Siegel, 1989), the surrounding debate is currently being boosted by the increased acknowledgment of the shortcomings associated with the categorical conceptualization.

Using a construct validation approach, our study provides further support for a dimensional conceptualization of reading disorder, as we did not find any structural differences with respect to the phonological core deficit of reading disorder. Likewise, the distinctiveness hypothesis is challenged by a range of studies comparing reading disorder to other forms of poor reading: For instance, despite intensive research over the past decades, there is no convincing evidence that children with reading disorder differ from non IQ-discrepant poor readers with respect to the symptoms (Hoskyn & Swanson, 2000; Stuebing et al., 2002), the cognitive causes (e.g., Jiménez, Siegel, & López, 2003; Maehler & Schuchardt, 2011; Stuebing et al., 2002; Toth & Siegel, 1994), the genetic or neuroanatomical correlates (see Stanovich, 1994b, for a review), the response to intervention (Stage, Abbott, Jenkins, & Berninger, 2003) or the general course of their reading development (e.g., Flowers, Meyer, Lovato, Wood, & Felton, 2000; O'Malley, Francis, Foorman, Fletcher, & Swank, 2003). Together, those findings are commonly interpreted as evidence for the low validity of current IQ-discrepancy models in particular and the categorical approach in general (e.g., Siegel, 1992; Stanovich, 1994b; Stuebing et al., 2002).

In addition, a range of methodological papers have pointed to the negative measurement issues that result from the dichotomization of continuously distributed reading and IQ scores. For example, the arbitrary nature of the thresholds as well as problems due to regression to the mean are well-recognized phenomena (e.g., Francis et al., 2005; Sternberg & Grigorenko, 2002). The low diagnostic stability of current RD classification further emphasizes the limitations of the categorical approach (e.g., Brown Waesche, Schatschneider, Maner, Ahmed, & Wagner, 2011; Francis et al., 2005; Schatschneider, Wagner, Hart, & Tighe, 2016). Moreover, splitting the reading and IQ continuum produces an information loss within the established categories and may also create statistical artefacts and may reduce statistical power (e.g., Cohen, 1983; McCallum, Zhang, Preacher, & Rucker, 2002; Maxwell & Delaney, 1993).

To summarize, current research points to the conclusion that a dimensional conceptualization of reading disorder might be more appropriate than a strictly categorical definition. Particularly, by acknowledging the quantitative nature of reading disorder we might be able to improve diagnostic validity. Further, there is justified hope that a dimensional approach would significantly enhance diagnostic reliability, which would help to overcome methodological shortcomings of current diagnostics. However, further studies would still have to determine the diagnostic criteria for a dimensional assessment of reading disorder. Overall, striving for a more valid and reliable classification of reading disorder remains a major challenge and is of utmost importance for the improvement of intervention planning for affected children.

Limitations of the Study and Directions for Further Studies

Although our study definitely contributes to the literature on phonological processing in reading disorder, there are some limitations worthwhile to be considered in further research. First, although our RD group was carefully selected as part of an extreme group design and although our cut-off scores followed the diagnostic guidelines commonly used in German educational practice, thresholds to define reading disorder are to some extent arbitrary. It might thus be possible that the tripartite structure does not hold in children who experience even more severe learning problems than the children participating in our study. Further, reading disorder is typically conceptualized categorically. However, there is no international agreement on how this category is defined: some countries such as Germany adhere to ICD-10 (WHO, 2011), others to DSM-5 (APA, 2013); there are also country-specific regulations. That is, countries differ in their concept of a reading disorder, and consequently different categorical groups are formed. Against this backdrop, it has to be kept in mind that results from our study may not necessarily transfer to countries in which, for example, DSM-5 criteria are applied in RD diagnostics.

Second, our analysis is based on a limited number of tasks. For instance, we assessed PA only on the phoneme level and here mainly with vowel sounds. Future studies may therefore also incorporate subtests that refer to rhymes and syllables as well as subtests that emphasize consonant sounds. In addition, some of the subtests measuring the same underlying construct showed only low correlations and in line with this finding some of the standardized factor loadings were only in the moderate range with values of .35 to .44. Although all these factor loadings were significant, we suggest that future studies should replicate our study with measures that are related to each other to a greater extent. Likewise, the sample size of 209 children is relatively small for a multi-group analysis of invariance testing. Small samples may reduce the power to differentiate between competing models and may also produce parameter estimates with large confidence intervals. Although similar sample sizes were also used in previous studies examining measurement invariance in child mental disorders (e.g., attention deficit hyperactivity disorder: Karalunas, Bierman, & Huang-Pollock, 2016; intellectual disabilities: Marsh, Tracey, & Craven, 2006; learning disabilities:

Schuchardt et al., 2008), future studies should replicate the present findings based on larger sample sizes.

Third, we examined the structural organization of phonological processing at a relatively late point in children's phonological development. The rationale for investigating the nature of phonological skills among third graders was that this is the age group for which reading disorder is most frequently diagnosed (cf. Hasselhorn & Schuchardt, 2006). Accordingly, a vast amount of studies conducted in the field targets this age group when analyzing performance differences between children with reading disorder and typical learners, which further justifies choosing this age group as a starting point for empirical examinations of invariance. Nevertheless, we suggest that it could be worthwhile to perform a similar analysis with kindergarten children *at risk* of reading disorder: Possibly, structural differences in phonological processing might exist at an earlier point in children's development. Likewise, it remains to be seen whether invariance would also hold longitudinally across grade levels: Since phonological processing is reciprocally related to emergent literacy skills (e.g., Burgess & Lonigan, 1998; Chow, McBride-Chang, & Burgess, 2005), it cannot be taken for granted that the structure found in one grade level automatically transfers to another grade level. Thus, longitudinal studies would help dertermine the structural *development* as well as the *longitudinal* invariance of phonological processing in children with reading disorder during the first years of formal reading instruction and beyond.

Last but not least, future studies could further expand our theoretical understanding concerning the structure of phonological processing by comparing different subgroups of reading disability. For instance, it would be of interest whether children whose reading problems are manifest mainly in the domain of word decoding (referred to as *dyslexia* in the Simple View of Reading; Gough & Tunmer, 1986) show the same or a different phonological processing structure as children whose problems concentrate mainly on reading comprehension and who exhibit a critical discrepancy between their word identification and text comprehension skills. Given the current debate as to whether these different phenotypes of reading disorder are associated with the same or distinct cognitive deficits (see Snowling and Hulme, 2012, for a review), those studies would be highly informative with respect to the question whether we are dealing with qualitatively different subgroups or whether the structure of phonological processing is transferable to different forms of reading disability.

Implications for Educational Practice

At least two educational implications can be drawn from our study, the first of which refers to diagnostics: When it comes to diagnosing reading disorder or to identifying at-risk children, an assessment of phonological processing is often part of the diagnostic procedure in order to better understand the child's strengths and weaknesses in the phonological domain. Our finding that common measures of phonological processing work equally well across typical learners and children with reading disorder implies that the results obtained from such a diagnostic assessment have the same underlying meaning in both groups and can thus be interpreted in the same way. If, however, the measurement instruments were not invariant, the test results would imply something different with respect to the underlying phonological constructs depending on the group a child is allocated to. Moreover, as Millsap and Oi-Man (2004) demonstrated, diagnostic decision-making may even be biased towards one of the groups: Specifically, missing group invariance may negatively influence the sensitivity of the instrument and may result in different selection and error rates depending on the subgroup. It is thus a crucial aspect of test fairness to know whether or not a measurement instrument works equally well across the subgroups for which it is used.

Finally, given the important role of phonological processing for emergent literacy, there is growing interest in fostering these skills through intervention–an approach that has as yet proven particularly effective for PA. As regards those phonological trainings, our findings may hold specific expectations with respect to cognitive transfer. In particular, the factor structure and interrelations we found may suggest that a training of PA, for instance, may also foster phonetic recoding in the PL (and vice versa), whereas cognitive transfer to RAN does not seem very promising. Likewise, it is unlikely to expect a training of RAN to transfer to PA and PL. Although the vast majority of intervention studies have not targeted those potential transfer effects, the few existing studies provide some support for this assumption. For instance, Regtvoort and van der Leij (2007) did not find a PA training to transfer to lexical access. In contrast, there is evidence that a training of PA may booster the storage capacity of the PL (Gillam, Kleeck, & Hoffman, 2006).

To summarize, this study contributed to our theoretical understanding of phonological processing by demonstrating factorial and measurement invariance of phonological processing skills between children with reading disorder and typical learners. In so doing, the present study closed a remaining significant gap in the RD literature.

Footnote

[1] Note that the general results of the second research question (i.e., testing for measurement invariance) were the same whether the correlated residual between object and color naming was included in the baseline model or not. In addition, we reran the analyses by using an item parcel that combined the object and color naming tasks into a single indicator in both groups rather than using on correlated residuals. Again, the results of the subsequent invariance tests were the same.

References

Authors (2013).

American Psychiatric Association. (2013). *Diagnostic and statistical manual of mental disorders – DSM-5* (5th ed.). Washington, DC: Author.

Anthony, J. L., & Lonigan, C. J. (2004). The nature of phonological awareness: Converging evidence from four studies of preschool and early grade school children. *Journal of Educational Psychology, 96*, 43–55. doi:10.1037/0022-0663.96.1.43

Anthony, J. L., Williams, J. M., McDonald, R., Corbitt-Shindler, D., Carlson, C. D., & Francis, D. J. (2006). Phonological processing and emergent literacy in Spanish-speaking preschool children. *Annals of Dyslexia, 56*, 239–270. doi:10.1007/s11881-006-0011-5

Anthony, J. L., Williams, J. M., McDonald, R., & Francis, D. J. (2007). Phonological processing and emergent literacy in younger and older preschool children. *Annals of Dyslexia, 57*, 113–137. doi:10.1007/s11881-007-0008-8

Archibald, L. M. D., & Gathercole, S. E. (2007). Nonword repetition and serial recall: Equivalent measures of verbal short-term memory? *Applied Psycholinguistics, 28*, 587–606. doi:10.1017/S0142716407070324

Aro, M., & Wimmer, H. (2003). Learning to read: English in comparison to six more regular orthographies. *Applied Psycholinguistics, 24*, 621–635. doi:10.1017.S0142716403000316

Babayiğit, S., & Stainthorp, R. (2011). Modeling the relationships between cognitive-linguistic skills and literacy skills: New insights from a transparent orthography. *Journal of Educational Psychology, 103*, 169–189. doi:10.1037/a0021671

Baddeley, A. D. (1986). *Working memory*. Oxford, UK: Oxford University Press.

Birkel, P. (2007). *Weingartener Grundwortschatz Rechtschreib-Test für zweite und dritte Klassen – WRT 2+* [Weingarten's spelling test of basic vocabulary for second and third grade]. Göttingen, Germany: Hogrefe.

Blomert, L., & Vaessen, A. (2008). *3DM differentiaal diagnostiek van dyslexie: Een cognitieve analyse van lezen en spellen* [3DM Differential diagnosis of dyslexia: A cognitive analysis of reading and spelling]. Amsterdam, Holland: Boom.

Boscardin, C. K., Muthén, B., Francis, D. J., & Baker, E. L. (2008). Early identification of reading difficulties using heterogeneous developmental trajectories. *Journal of Educational Psychology, 100*, 192–208. doi:10.1037/0022-0663.100.1.192

Bowers, P. G., & Wolf, M. (1999). The double-deficit hypothesis for the developmental dyslexias. *Journal of Educational Psychology, 91*, 415–438. doi:10.1037/0022-0663.91.3.415

Bowey, J. A. (1997). What does nonword repetition measure? A reply to Gathercole and Baddeley. *Journal of Experimental Child Psychology, 67*, 295–301. doi:10.1006/jecp.1997.2408

Bradley, L. & Byrant, P. E. (1985). *Rhyme and reason in reading and spelling*. Ann Arbor: University of Michigan Press.

Branum-Martin, L., Fletcher, J. M., & Stuebing, K. K. (2012). Classification and identification of reading and math disabilities: The case of comorbidity. *Journal of Learning Disabilities, 46*, 490–499. doi:10.1177/0022219412468767

Brown Waesche, J. S., Schatschneider, C., Maner, J. K., Ahmed, Y., & Wagner, R. K. (2011). Examing agreement and longitudinal stability among traditional and RTI-based definitions of reading disability using the affected-status agreement statistic. *Journal of Learning Disabilities, 44*, 296–307. doi:10.1177/0022219410392048

Burgess, S. R. & Lonigan, C. J. (1998). Bidirectional relations of phonological sensitivity and pre-reading abilities: Evidence from a preschool sample. *Journal of Experimental Child Psychology, 70*, 117–141. doi:10.1006/jecp.1998.2450

Byrne, B. M. (2012). *Structural equation modeling with Mplus.* New York, NY: Routledge.

Cattell, R. B., Weiß, R. H., & Osterland, J. (1997). *Grundintelligenztest Skala 1 – CFT 1* [Basic intelligence test scale 1] (5th ed.). Göttingen, Germany: Hogrefe.

Cheung, G. W., & Rensvold, R. B. (2002). Evaluating goodness-of fit indexes for testing measurement invariance. *Structural Equation Modeling: A Multidisciplinary Journal, 9*, 233–255. doi:10.1207/S15328007SEM0902_5

Chow, B.W.-Y. , McBride-Chang, C., & Burgess, S. (2005). Phonological processing skills and early reading abilities in Hong Kong Chinese kindergarteners learning to read English as a second language. *Journal of Educational Psychology*, 97, 81–87. doi: 10.1037/0022-0663.97.1.81

Cohen, J. (1983). The cost of dichotomization. *Applied Psychological Measurement, 7*, 249–253. doi:10.1177/014662168300700301

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). New York, NY: Academic Press.

Coghill, D., & Sonuga-Barke, E. J. S. (2012). Categories versus dimensions in the classification and conceptualisation of child and adolescent mental disorders implications of recent empirical study. *The Journal of Child Psychology and Psychiatry, 53*, 469–489. doi:10.1111/j.1469-7610.2011.02511.x

Compton, D. L., Fuchs, L. S., Fuchs, D., Lambert, W., & Hamlett, C. (2012). The cognitive and academic profiles of reading and mathematics learning disabilities. *Journal of Learning Disabilities, 45*, 79–95. doi:10.1177/0022219410393012 de Jong, P. F., & van der Leij, A. (1999). Specific contributions of phonological abilities to early reading acquisition: Results from a Dutch latent variable longitudinal study. *Journal of Educational Psychology, 91*, 450–476. doi:10.1037/0022-0663.91.3.450

Denckla, M. B., & Rudel, R. G. (1976). Naming of object-drawings by dyslexic and other learning disabled children. *Brain and Language, 3*, 1–15. doi:10.1016/0093-934X(76)90001-8

Dirks, E., Spyer, G., van Lieshout, E. C. D. M., & de Sonneville, L. (2008). Prevalence of combined reading and arithmetic disabilities. *Journal of Learning Disabilities, 41*, 460–473. doi:10.1177/0022219408321128

Ennemoser, M., Marx, P., Weber, J., & Schneider, W. (2012). Spezifische Vorläuferfertigkeiten der Lesegeschwindigkeit, des Leseverständnisses und des Rechtschreibens: Evidenz aus zwei Längsschnittstudien vom Kindergarten bis zur 4. Klasse [Specific precursors of decoding speed, reading comprehension, and spelling: Evidence from two longitudinal studies from kindergarten to grade 4]. *Zeitschrift für Entwicklungspsychologie und Pädagogische Psychologie, 44*, 53–67. doi:10.1026/0049-8637/a000057

Ferrer, E., Shaywitz, B. A., Hoahan, J. M., Marchione, K., & Shaywitz, S. E. (2010). Uncoupling of reading and IQ over time: Empirical evidence for a definition of dyslexia. *Psychological Science*, *21*, 93–101. doi:10.1177/0956797609354084

Flowers, L., Meyer, M., Lovato, J., Wood, F. & Felton, R. (2000). Does third grade discrepancy status predict the course of reading development? *Annals of Dyslexia, 50*, 49–71. doi:10.1007/s11881-001-0005-2

Francis, D. J., Fletcher, J. M., Stuebing, K. K., Lyon, G. R., Shaywitz, B. A., & Shaywitz, S. E. (2005). Psychometric approaches to the identification of LD: IQ and achievement scores are not sufficient. *Journal of Learning Disabilities, 38*, 98–108. doi:10.1177/00222194050380020101

Furnes, B., & Samuelsson, S. (2011). Phonological awareness and rapid automatized naming predicting early development in reading and spelling: Results from a cross-linguistic longitudinal study. *Learning and Individual Differences, 21*, 85–95. doi:10.1016/j.lindif.2010.10.005

Gathercole, S. E. (2006). Nonword repetition and word learning: The nature of the relationship. *Applied Psycholinguistics, 27*, 513–543. doi:10.1017.S0142716406060383

Gathercole, S. E., & Baddeley, A. D. (1993). *Working memory and language*. Hove: Erlbaum.

Gillam, R. B., Kleeck, A. V., & Hoffman, L. M. (2006). Training in phonological awareness generalizes to phonological working memory: A preliminary investigation. *The Journal of Speech and Language Pathology – Applied Behavior Analysis, 1*, 228–243. doi: 10.1037/h0100201

Gough, P. B., & Tunmer, W. E. (1986). Decoding, reading and reading disability. *Remedial and Special Education, 7*, 6–10. doi:10.1177/074193258600700104

Hasselhorn, M., Grube, D., & Mähler, C. (2000). Theoretisches Rahmenmodell für ein Diagnostikum zur differentiellen Funktionsanalyse des phonologischen Arbeitsgedächtnisses [Theoretical framework for diagnosing the differential functional analysis of phonological working memory]. In M. Hasselhorn, W. Schneider & H. Marx (Eds.), *Diagnostik von Lese-Rechtschreibschwierigkeiten* (Test und Trends N. F., Bd. 1. Jahrbuch der pädagogisch-psychologischen Diagnostik, pp. 167–182). Göttingen, Germany: Hogrefe.

Hasselhorn, M., & Schuchardt, K. (2006). Lernstörungen: Eine kritische Skizze zur Epidemiologie [Learning disabilities: A critical sketch on epidemiology]. *Kindheit und Entwicklung, 15*, 208–215. doi:10.1026/0942-5403.15.4.208

Hasselhorn, M., Schumann-Hengsteler, R., Grube, D., König, J., Mähler, C., Schmid, I., … Zoelch, C. (2012). *Arbeitsgedächtnistestbatterie für Kinder von 5 bis 12 Jahren (AGTB 5–12)* [Working memory test battery for children aged five to twelve years]. Göttingen, Germany: Hogrefe.

Hoskyn, M. & Swanson, H. L (2000). Cognitive processing of low achievers and children with reading disabilities: A selective meta-analytic review of the published literature. *The School Psychology Review, 29*, 102–119.

Hu, L., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling: A Multidisciplinary Journal, 6*, 1–55. doi:10.1080/10705519909540118

Jiménez, J. E., Siegel, L. S., & Rodrigo López, M. (2003). The relationship between IQ and reading disabilities in English speaking Canadian and Spanish children. *Journal of Learning Disabilities, 36*, 15–23. doi:10.1177/00222194030360010301

Karalunas, S. L., Bierman, K. L., & Huang-Pollock, C. L. (2016). Test-retest reliability and measurement invariance of executive function tasks in young children with and without ADHD. *Journal of Attention Disorder.* Advance online publication. doi:10.1177/1087054715627488

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*(3), 341–362. doi:10.1598/RRQ.45.3.4

Kline, R. B. (2016). *Principles and practice of structural equation modeling* (4th ed.). New York, NY: The Guilford Press.

Krajewski, K., Liehm, S., & Schneider, W. (2004). *Deutscher Mathematiktest für zweite Klassen (DEMAT 2+)* [German test of mathematical skills for second grades]. Göttingen, Germany: Beltz.

Kudo, M. F., Lussier, C. M., & Swanson, H. L. (2015). Reading disabilities in children: A selective meta-analysis of the cognitive literature. *Research in Developmental Disabilities, 40*, 51–62. doi:10.1016/jrid.d.2015.01.002

Lambrecht Smith, S., Scott, K. A., Roberts, J., & Locke, J. L. (2008). Disabled readers' performance on tasks of phonological processing, rapid naming, and letter knowledge before and after kindergarten. *Learning Disabilities Research & Practice, 23*, 113–124. doi:10.1111/j.1540-5826.2008.00269.x

Landerl, K. (2001). Word recognition deficits in German: More evidence from a representative sample. *Dyslexia*, *7*, 183–196. doi:10.1002/dys.199

Landerl, K. (2003). Categorization of vowel length in German poor spellers: An orthographically relevant phonological distinction. *Applied Psycholinguistics, 24*, 523–538. doi:10.1017.S0142716403000262

Landerl, K., & Moll, K. (2010). Comorbidity of learning disorders: prevalence and familial transmission. *Journal of Child Psychology and Psychiatry*, *51*, 287–294. doi:10.1111/j.1469-7610.2009.02164.x

Landerl, K., Wimmer, H., & Frith, U. (1997). The impact of orthographic consistency on dyslexia: A German-English comparison. *Cognition*, *63*, 315–334. doi:10.1016/S0010-0277(97)00005-X

Lenhard, W., & Schneider, W. (2006). *Ein Leseverständnistest für Erst- bis Sechstklässler (ELFE 1–6)* [A reading comprehension test for first to sixth graders]. Göttingen, Germany: Hogrefe.

Logan, J. A. R., Schatschneider, C., & Wagner, R. K. (2011). Rapid serial naming and reading ability: The role of lexical access. *Reading and Writing: An Interdisciplinary Journal, 24*, 1–15. doi:10.1007/s11145-009-9199-1

Lonigan, C. J., Anthony, J. L., Philips, B. M., Purpura, D. J., Wilson, S. B., & McQueen, J. D. (2009). The nature of preschool phonological processing abilities and their relations to vocabulary, general cognitive abilities, and print knowledge. *Journal of Educational Psychology, 101*, 345–358. doi:10.1037/a0013837

Lyon, G. R., Shaywitz, S. E., & Shaywitz, B. A. (2003). A definition of dyslexia. *Annals of Dyslexia, 53*, 1–13. doi:10.1007/s11881-003-0001-9

MacCallum, R. C., Zhang, S., Preacher, K. J., & Rucker, D. D. (2002). On the practice of dichotomization of quantitative variables. *Psychological Methods, 7*, 19–40. doi:10.1037//1082-989X.7.1.19

Maehler, C., & Schuchardt, K. (2011). Working memory in children with learning disabilities: Rethinking the criterion of discrepancy. *International Journal of Disability, Development and Education, 58*, 5–17. doi:10.1080/1034912X.2011.547335

Mardia, K. V. (1974). Applications of some measures of multivariate skewness and kurtosis in testing normality and robustness studies. *Sankhyā: The Indian Journal of Statistics – Series B, 36*, 115–128.

Marsh, H. W., Tracey, D. K., & Craven, G. H. (2006). Multidimensional self-concept structure for preadolescents with mild intellectual disabilities: A hybrid multigroup–MIMC approach to factorial invariance and latent mean differences. *Educational and Psychological Measurement, 66*, 795–818. doi: 10.1177/0013164405285910.

Maxwell, S. E., & Delaney, H. D. (1993). Bivariate median splits and spurious statistical significance. *Quantitative Methods in Psychology, 113*, 181–190. doi:10.1037/0033-2909.113.1.181

Melby-Lervåg, M., Lyster, S.-A. H., & Hulme, C. (2012). Phonological skills and their role in learning to read: A meta-analytic review. *Psychological Bulletin, 138*, 322–352. doi:10.1037/a0026744

Meredith, W. (1993). Measurement invariance, factor analysis and factorial invariance. *Psychometrika, 58*, 525–543. doi:10.1007/BF02294825

Metsala, J. L. (1999). Young children's phonological awareness and nonword repetition as a function of vocabulary development. *Journal of Educational Psychology, 91*, 3–19. doi:10.1037/0022-0663.91.1.3

Meyer, M. S. (2000). The ability–achievement discrepancy: Does it contribute to an understanding of learning disabilities? *Educational Psychology Review, 12*, 315–337. doi:10.1023/A:1009070006373

Michalczyk, K., Malstädt, N., Worgt, M., Könen, T. & Hasselhorn, M. (2013). Age differences and measurement invariance of working memory in 5- to 12-year-old children. *European Journal of Psychological Assessment, 29*, 220–229. doi:10.1027/1015-5759/a000149

Millsap, R.E. (2011). *Statistical approaches to measurement invariance*. New York: Routledge.

Millsap, R. E., & Oi-Man, K. (2004). Evaluating the impact of partial factorial invariance on selection in two populations. *Psychological Methods, 9*, 93–115. doi:10.1037/1082-989X.9.1.93

Mody, M. (2003). Phonological basis in reading disability: A review and analysis of the evidence. *Reading and Writing: An Interdisciplinary Journal, 16*, 21–39. doi:10.1023/A:1021741921815

Moll, K., Bruder, J., Kunze, S., Neuhoff, N., & Schulte-Körne, G. (2014). Specific learning disorder: Prevalence and gender differences. *Plos One, 9*, 1–8. doi:10.1371/journal.pone.0103537

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), 1–25. doi:10.1080/10888430802631684

Moll, K., Ramus, F., Bartling, J., Bruder, J., Kunze, S., Neuhoff, N. et al. (2014). Cognitive mechanisms underlying reading and spelling development in five European orthographies. *Learning and Instruction, 29*, 65–77. doi:10.1016/j.learninstruc.2013.09.003

Muthén, L. K., & Muthén, B. O. (1998-2012). *Mplus user's guide* (7th ed.). Los Angeles, CA: Muthén and Muthén.

Niedersächsisches Kultusministerium. (2006). *Kerncurriculum für die Grundschule. Schuljahrgänge 1-4*. [*core-curriculum for elementary school. Grades 1-4*]. Retrieved from http://db2.nibis.de/1db/cuvo/ausgabe/

Nikolopoulos, D., Goulandris, N., Hulme, C., & Snowling, M. J. (2006). The cognitive bases of learning to read and spell in Greek: Evidence from a longitudinal study. *Journal of Experimental Child Psychology, 94*, 1–17. doi:10.1016/j.jecp.2005.11.006

Norton, E. S., & Wolf, M. (2012). Rapid automatized naming (RAN) and reading fluency: Implications for understanding and treatment of reading disabilities. *Annual Review of Psychology, 63*, 427–452. doi:10.1146/annurev-psych-120710-100431

Nunnally, J. C., & Bernstein, I. H. (1994). *Psychometric theory* (3rd Ed.). New York, NY: McGraw-Hill.

O'Malley, K. J., Francis, D. J., Foorman, B. R., Fletcher, J. M., & Swank, P. R. (2002). Growth in precursor and reading-related skills: Do low-achieving and IQ-discrepant readers develop differently? *Learning Disabilities Research & Practice, 17*, 19–34. doi:10.1111/1540-5826.00029

Papadopoulos, T. C., Kendeou, P., & Spanoudis, G. (2012). Investigating the factor structure and measurement invariance of phonological abilities in a sufficiently transparent language. *Journal of Educational Psychology, 104*, 321–336. doi:10.1037/a0026446

Papadopoulos, T. C., Spanoudis, G., & Kendeou, P. (2009). The dimensionality of phonological abilities in Greek. *Reading Research Quarterly, 44*, 127–143. doi:10.1598/RRQ.44.2.2

Peterson, R. L., & Pennington, B. F. (2015). Developmental dyslexia. *Annual Review of Clinical Psychology, 11*, 283–307. doi:10.1146/annurev-clinpsy-032814-112842

Regtvoort, G. F. M., & van der Leij, A. (2007). Early intervention with children of dyslexic parents: Effects of computer-based reading instruction at home on literacy acquisition. *Learning and Individual Differences, 17*, 35–53. doi:10.1016/j.lindif.2007.01.005

Satorra, A., & Bentler, P. M. (2001). A scaled difference chi-square test statistic for moment structure analysis. *Psychometrica, 66*, 507–514. doi:10.1007/BF02296192

Schatschneider, C., Francis, D. J., Forrman, B. R., Fletcher, J. M., & Mehta, P. (1999). The dimensionality of phonological awareness: An application of item response theory. *Journal of Educational Psychology, 91*, 439–449. doi:10.1037/0022-0663.91.3.439

Schatschneider, C., Wagner, R. K., Hart, S. A., & Tighe, E. L. (2016). Using simulations to investigate the longitudinal stability of alterntive schemas for classifying and identifying children with reading disabilities. *Scientific Studies of Reading, 20*, 34–48. doi:https://dx.doi.org/10.1080/10888438.2015.1107072

Schneider, W., & Näslund, J. C. (1993). The impact of early metalinguistic competencies and memory capacity on reading and spelling in elementary school: Results of the Munich longitudinal study on the genesis of individual competencies (LOGIC). *European Journal of Psychology of Education, 8*, 273–287. doi:10.1007/BF03174082

Schuchardt, K., Roick, T., Mähler, C., & Hasselhorn, M. (2008). Unterscheidet sich die Struktur des Arbeitsgedächtnisses bei Schulkindern mit und ohne Lernstörung? [Does learning disability make a difference regarding the structure of working memory in children?] *Zeitschrift für Entwicklungspsychologie und Pädagogische Psychologie, 40*, 147–151. doi:10.1026/0049-8637.40.3.147

Siegel, L. S. (1989). IQ is irrelevant to the definition of learning disabilities. *Journal of Learning Disabilities, 22*, 469–478. doi:10.1177/002221948902200803

Siegel, L. S. (1992). An evaluation of the discrepancy definition of dyslexia. *Journal of Learning Disabilities, 25*, 618–629. doi:10.1177/002221949202501001

Smythe, I., Everatt, J., Al-Menaye, N., He, X., Capellini, S., Gyarmathy, E., & Siegel, L. S. (2008). Predictors of word-level literacy amongst Grade 3 children in five diverse languages. *Dyslexia, 14*, 170–187. doi:10.1002/dys.369

Snowling, M. J., & Hulme, C. (2012). The nature and classification of reading disorders – a commentary on proposals for DSM-5. *The Journal of Child Psychology and Psychiatry, 53*, 593–607. doi:10.1111/j.1469-7610.2011.02495.x

Sprugevica, I., & Høien, T. (2004). Relations between enabling skills and reading comprehension: A follow-up study of Latvian students from first to second grade. *Scandinavian Journal of Psychology, 45*, 115–122. doi:10.1111/j.1467-9450.2004.00386.x

Stage, S. A., Abbott, R. D., Jenkins, J. R., & Berninger, V. W. (2003). Predicting response to early reading intervention from verbal IQ, reading-related language abilities, attention ratings, and verbal IQ–word reading discrepancy: Failure to validate discrepancy method. *Journal of Learning Disabilities, 36*, 24–33. doi:10.1177/00222194030360010401

Stanovich, K. E. (1988). Explaining the differences between the dyslexic and the garden-variety poor reader: The phonological-core variable-difference model. *Journal of Learning Disabilities, 21*, 590–604. doi:10.1177/002221948802101003

Stanovich, K. E. (1994a). Does dyslexia exist? *Child Psychology & Psychiatry & Applied Disciplines, 35*, 579–595. doi:10.1111/j.1469-7610.1994.tb01208.x

Stanovich, K. E. (1994b). Are discrepancy-based definitions of dyslexia empirically defensible? In K. P. van den Bos, L. S. Siegel, D. J. Bakker, & D. L. Share, (Eds.), Current directions in dyslexia research (pp. 15–30). Lisse, Netherlands: Swets & Zeitlinger.

Sternberg, R. J., & Grigorenko, E. L. (2002). Difference scores in the identification of children with learning disabilities – It's time to use a different method. *Journal of School Psychology, 40*, 65–83. doi:10.1016/S0022-4405(01)00094-2

Stock, C., Marx, P., & Schneider, W. (2003). *Basiskompetenzen für Lese-Rechtschreibleistungen: Ein Test zur Erfassung der phonologischen Bewusstheit vom ersten bis vierten Schuljahr* (*BAKO 1–4)* [Basic competencies of reading and spelling skills: A test assessing phonological awareness from first to fourth grade]. Göttingen, Germany: Beltz.

Stuebing, K. K., Fletcher, J. M., LeDoux, J. M., Lyon, G. R., Shaywitz, S. E., & Shaywitz, B. A. (2002). Validity of IQ-discrepancy classifications of reading disabilities: A meta-analysis. *American Educational Research Journal*, *39*, 469–518. doi:10.3102/00028312039002469

Swanson, H. L. (2012). Adults with reading disabilities: Converting a meta-analysis to practice. *Journal of Learning Disabilities, 45*, 17–30. doi:10.1177/0022219411426856

Swanson, H. L., & Hsieh, C.-J. (2009). Reading disabilities in adults: A selective meta-analysis of the literature. *Review of Educational Research, 79*, 1362–1390. doi:10.3102/0034654309350931

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. doi:10.3102/00346543073004407

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: An Interdisciplinary Journal, 26*, 1352–1380. doi:10.1007/s11145-012-9423-2

Toth, G., & Siegel, L. S. (1994). A critical evaluation of the IQ-achievement discrepancy based definitions of dyslexia. In K. P. van den Bos, L. S. Siegel, D. J. Bakker, D. L. Share, (Eds.), Current directions in dyslexia research (pp. 45–70). Lisse, Netherlands: Swets & Zeitlinger.

Vaessen, A., Bertrand, D., Tóth, D., Csépe, V., Faísca, L., Reis, A., & Blomert, L. (2010). Cognitive development of fluent word reading does not qualitatively differ between transparent and opaque orthographies. *Journal of Educational Psychology, 102*, 827–842. doi:10.1037/a0019465

Vandenberg, R. J., & Lance, C. E. (2000). A review and synthesis of the measurement invariance literature: Suggestions, practices, and recommendations for organizational research. *Organizational Research Methods, 3*, 4–70. doi:10.1177/109442810031002

van de Schoot, R., Lugtig, P., & Hox, J. (2012). A checklist for testing measurement invariance. *European Journal of Developmental Psychology, 9*, 486–492. doi:10.1080/17405629.2012.686740

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. doi:10.1046/j.0021-9630.2003.00305.x

Vloedgraven, J., & Verhoeven, L. (2009). The nature of phonological awareness throughout the elementary grades: An item response theory perspective. *Learning and Individual Differences, 19*, 161–169. doi:10.1016/j.lindif.2008.09.005

Wagner, R. K. (1986). Phonological processing abilities and reading: Implications for disabled readers. *Journal of Learning Disabilities, 19*, 623–630. doi:10.1177/002221948601901009

Wagner, R. K., Balthazor, M., Hurley, S., Morgan, S., Rashotte, C., Shaner, R., …, Stage, S. (1987). The nature of prereaders' phonological processing abilities. *Cognitive Development, 2*, 355–373. doi:10.1177/002221948601901009

Wagner, R. K., & Torgesen, J. K. (1987). The nature of phonological processing and its causal role in the acquisition of reading skills. *Psychological Bulletin, 101*, 192–212. doi:10.1037//0033-2909.101.2.192

Wagner, R. K., Torgesen, J. K., Laughon, P., Simmons, K., & Rashotte, C. A. (1993). Development of young readers' phonological processing abilities. *Journal of Educational Psychology, 85*, 83–103. doi:10.1037/0022-0663.85.1.83

Wagner, R. K., Torgesen, J. K., & Rashotte, C. A. (1994). Development of reading-related phonological processing abilities: New evidence of bidirectional causality from a latent variable longitudinal study. *Developmental Psychology, 30*, 73–87. doi:10.1037//0012-1649.30.1.73

Wang, J., & Wang, X. (2012). *Structural equation modeling: Applications using Mplus*. West Sussex, UK: John Wiley & Sons.

Wimmer, H., Landerl, K., Linorter, R., & Hummer, P. (1991). The relationship of phonemic awareness to reading acquisition: More consequence than precondition but still important. *Cognition, 40*, 219–249. doi:10.1016/0010-0277(91)90026-Z

Wimmer, H., & Schurz, M. (2010). Dyslexia in regular orthographies: Manifestation and causation. *Dyslexia, 16*, 283–299. doi:10.1002/dys.411

Wolf, M., Goldberg O'Rourke, A., Gidney, C., Lovett, M., Cirino, P., & Morris, R. (2002). The second deficit: An investigation of the independence of phonological and naming-speed deficits in developmental dyslexia. *Reading and Writing: An Interdisciplinary Journal, 15*, 43–72. doi:10.1023/A:1013816320290

World Health Organization (2011). *ICD: Classification of mental and behavioural disorders: Clinical descriptions and diagnostic guidelines* (10th. rev. ed.). Geneva, Switzerland: Author.

Ziegler, J. C., Perry, C., Ma-Wyatt, A., Ladner, D., & Schulte-Körne, G. (2003). Developmental dyslexia in different languages: Language-specific or universal? *Journal of Experimental Child Psychology, 86*, 169–193. doi:10.1016/S0022-0965(03)00139-5

Table 1

Descriptive Statistics for All Measures as a Function of Group

	Reference Group (*n* = 100)				RD Group (*n* = 109)			
Measures	*M*	*SD*	Skew.	Kurt.	*M*	*SD*	Skew.	Kurt.
	Classification Measures (Independent Variables) and Age							
Age (in month)	102.16	4.86	-0.43	-0.21	103.36	5.80	0.60	0.02
Intelligence[a]	106.86	11.19	0.77	0.52	108.07	9.68	0.96	1.17
Mathematics[b]	53.98	5.49	0.42	-0.43	53.40	6.53	0.06	-0.83
Reading[b]	53.53	5.91	0.95	0.99	39.86	8.03	0.75	-0.46
Spelling[b]	51.36	5.75	0.92	0.27	37.77	6.31	0.58	-0.01
	Phonological Processing Measures (Dependent Variables)							
Phon. Awareness								
Vowel Length	4.63[c]	2.76	-0.05	-0.95	3.45[c]	2.25	0.31	-0.58
Vowel Substitution	9.46[d]	2.28	-1.04	0.76	7.06[c]	2.94	-0.51	-0.37
Phoneme Reversal	9.41[c]	4.79	-0.04	-1.25	5.21[d]	4.16	0.84	0.05
Phonological Loop								
1-syll. Word Span	3.95	0.65	0.08	-0.02	3.74	0.58	-0.27	-0.52
3-syll. Word Span	3.12	0.44	0.37	0.45	3.00[d]	0.37	0.17	0.30
Digit Span	4.61	0.60	-0.41	0.70	4.17	0.58	-0.36	-0.14
Serial Naming (in s)								
Color Naming	49.93	10.20	0.78	0.81	51.86[e]	10.56	0.82	0.33
Digit Naming	28.88[c]	6.20	0.94	1.04	34.25[f]	7.29	0.62	-0.06
Letter Naming	31.30	6.87	1.16	1.34	34.35[g]	7.18	1.10	1.72
Object Naming	47.42[d]	7.53	0.54	0.23	48.29[f]	7.41	0.63	0.96

Note. RD = Reading Disorder; Skew. = skewness; Kurt. = kurtosis; Phon. = phonological; syll. = syllable.

[a] IQ-score; [b] *T*-score; [c] Data of 1 participant are missing; [d] Data of 2 participants are missing; [e] Data of 3 participants are missing; [f] Data of 5 participants are missing; [g] Data of 4 participants are missing.

Overall, less than 1.5% of the phonological processing data was missing. There were various reasons for missing data, which can best be summarized as technical problems or test administration errors that occurred during the testing (e.g., children did not understand the test correctly or did not complete the task, technical problems with data recording especially during the RAN test, student research assistants did not administer the test correctly).

Table 2

Correlations among Phonological Processing Measures

Measures	1	2	3	4	5	6	7	8	9	10
1. Vowel Substitution	-	.14	.30	.21	.22	.14	.11	.16	.16	.18
2. Vowel Length	.19	-	.39	.25	.26	.27	-.16	-.10	-.08	.00
3. Phoneme Reversal	.40	.22	-	.06	.15	.09	-.06	.06	.03	.04
4. 1-syllabic Word Span	.21	.08	.10	-	.63	.65	.05	-.13	-.10	-.08
5. 3-syllabic Word Span	.20	.22	.28	.47	-	.67	.07	-.02	.03	.01
6. Digit Span	.20	.17	.17	.57	.47	-	-.04	-.13	-.05	-.03
7. Object Naming	-.13	-.10	-.03	-.21	-.09	-.07	-	.60	.29	.35
8. Color Naming	-.09	-.01	-.06	-.06	-.04	-.08	.50	-	.36	.42
9. Digit Naming	.04	.01	-.08	-.12	.07	-.08	.39	.47	-	.57
10. Letter Naming	-.07	-.16	-.13	-.02	.10	-.04	.25	.29	.44	-

Note. Intercorrelations for the Reference Group are presented above, the ones for the RD Group are presented below the diagonal.

All correlation coefficients of $r \geq .20$ are significant at $p < .05$.

Table 3

Measurement Model of Phonological Processing as a Function of Group

	Reference Group							RD Group						
CFA models	χ^2	*df*	*p*	CFI	RMSEA [90%CI]	AIC	BIC	χ^2	*df*	*p*	CFI	RMSEA [90%CI]	AIC	BIC
Two-Factor	65.75	34	<.001	.87	.10 [.06, .13]	4566.79	4647.55	56.25	34	.010	.88	.08 [.04, .11]	4909.07	4992.50
Three-Factor	54.16	32	.009	.91	.08 [.04, .12]	4555.84	4641.81	37.84	32	.220	.97	.04 [.00,. 09]	4896.05	4984.87
Three-Factor[a]	30.80	31	.476	1.00	.00 [.00, .07]	4539.51	4628.09	-	-	-	-	-	-	-

Note. RD = Reading Disorder; CFI = comparative fit index; RMSEA = root mean square error of approximation; AIC = Akaike's information criterion; BIC = Bayesian information criterion.

[a] Three-factor model with correlated residual between the object naming task and the color naming task, as indicated by modification indices.

Table 4

Factor Loadings and Item Uniquenesses of the Phonological Processing Model

	RD Group				Reference Group			
	Loadings		Uniquenesses		Loadings		Uniquenesses	
	Est. (S.E.)	Std. (S.E.)	Est. (S.E.)	Std. (S.E.)	Est. (S.E.)	Std. (S.E.)	Est. (S.E.)	Std. (S.E.)
Rapid Automatized Naming								
Object Naming	1.000 (0.000)	0.641 (0.093)	32.480 (6.983)	0.590 (0.119)	1.000 (0.000)	0.435 (0.111)	45.390 (8.196)	0.811 (0.096)
Colour Naming	1.541 (0.287)	0.697 (0.094)	56.874 (16.601)	0.514 (0.131)	1.646 (0.381)	0.527 (0.097)	74.365 (17.161)	0.722 (0.102)
Digit Naming	1.078 (0.325)	0.703 (0.106)	26.94 (7.862)	0.506 (0.149)	1.375 (0.450)	0.719 (0.080)	18.721 (4.454)	0.484 (0.115)
Letter Naming	0.746 (0.268)	0.496 (0.123)	38.550 (10.774)	0.754 (0.122)	1.680 (0.546)	0.799 (0.088)	16.854 (6.481)	0.361 (0.141)
Phonological Awareness								
Vowel Substitution	1.000 (0.000)	0.629 (0.115)	5.187 (1.240)	0.605 (0.144)	1.000 (0.000)	0.418 (0.192)	4.251 (1.238)	0.825 (0.161)
Vowel Length	0.424 (0.192)	0.349 (0.116)	4.398 (0.636)	0.878 (0.081)	1.685 (1.168)	0.584 (0.170)	4.960 (1.378)	0.659 (0.198)
Phoneme Reversal	1.379 (0.479)	0.614 (0.123)	10.633 (2.589)	0.623 (0.151)	3.082 (1.339)	0.614 (0.141)	14.145 (3.967)	0.623 (0.173)
Phonological Loop								
1-syllabic Word Span	1.000 (0.000)	0.737 (0.073)	0.154 (0.034)	0.456 (0.108)	1.000 (0.000)	0.786 (0.052)	0.158 (0.029)	0.383 (0.081)
3-syllabic Word Span	0.544 (0.133)	0.637 (0.082)	0.080 (0.013)	0.595 (0.105)	0.706 (0.109)	0.808 (0.050)	0.068 (0.015)	0.347 (0.080)
Digit Span	1.010 (0.173)	0.755 (0.070)	0.141 (0.033)	0.430 (0.106)	0.985 (0.131)	0.828 (0.048)	0.114 (0.027)	0.314 (0.079)

Note. RD = Reading Disorder; Est. = unstandardized parameter estimates; S.E. = standard error of the estimate; Std. = standardized parameter estimates.

Table 5

Factorial and Measurement Invariance of Phonological Processing: Fit Indices for Nested Model Comparison

	χ^2	*df*	p_1	Compared with	Δ SB-χ^2	Δ *df*	p_2	CFI	Δ CFI	RMSEA [90%CI]	AIC	BIC
Model 1	68.62	63	.293					.98		.03 [.00, .07]	9435.57	9659.50
Model 2	80.39	70	.186	Model 1	11.19	7	.130	.97	.01	.04 [.00, .07]	9433.88	9634.42
Model 3	96.13	80	.106	Model 2	14.99	10	.133	.96	.01	.04 [.00, .07]	9431.65	9598.76
Model 4	100.13	83	.097	Model 3	3.99	3	.263	.96	.00	.04 [.00, .07]	9429.77	9586.87
Model 5	102.17	86	.113	Model 4	2.13	3	.546	.96	.00	.04 [.00, .07]	9426.00	9573.06

Note. Model 1 = configural invariance (all parameters estimated freely); Model 2 = metric invariance (factor loadings constrained equal); Model 3 = strict invariance (factor loadings and residual variances constrained equal); Model 4 = invariance of factor variances (factor loadings, residual variances, and factor variances constrained equal); Model 5 = invariance of factor covariances (factor loadings, residual variances, factor variances and factor covariances constrained equal). CFI = comparative fit index; RMSEA = root mean square error of approximation; AIC = Akaike's information criterion; BIC = Bayesian information criterion. p_1 = probability value of model fit; p_2 = probability value obtained in the SB-χ^2–difference test. All models are estimated with the Robust Maximum Likelihood (MLR) estimator. In an additional analysis, Model 3 to 5 were compared with the baseline model (Model 1) rather than being tested against the respective previous model. Even under this stricter form of nested model comparison, all the resulting difference tests were nonsignificant – further confirming that measurement invariance holds across the two groups.

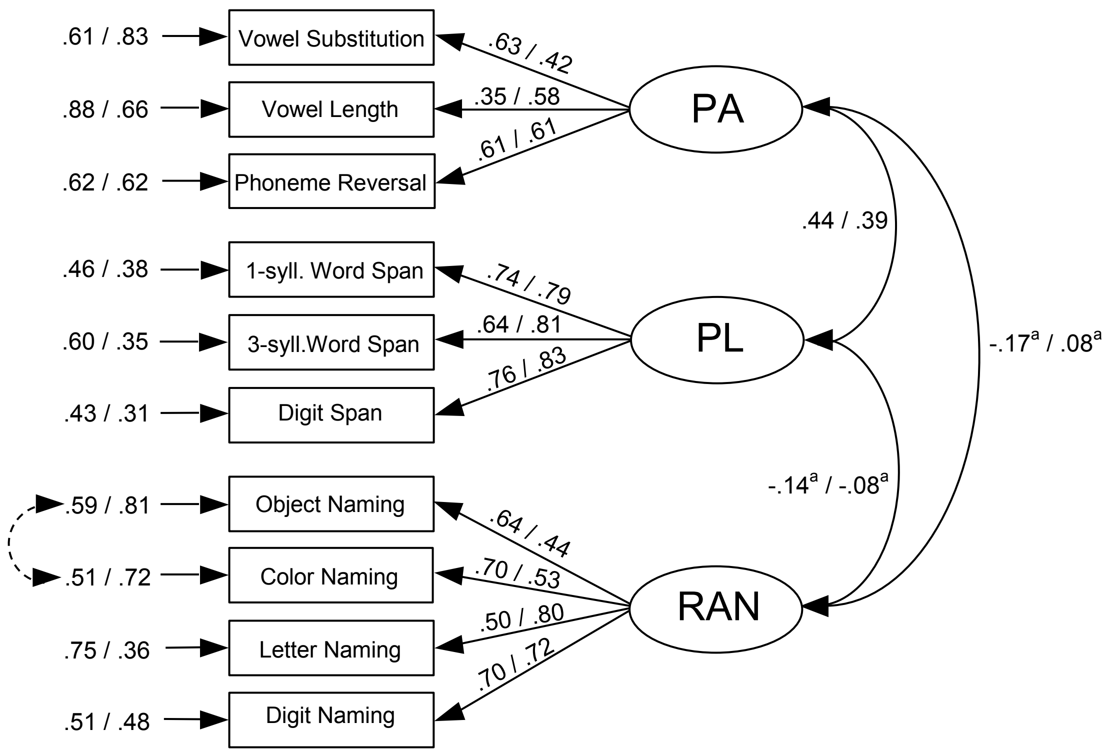

Figure 1. Three-factor oblique confirmatory model of phonological processing for children with specific reading disorder (left parameters) versus for children of the reference group (right parameters). Parameters represent standardized estimates. PA = phonological awareness; PL = phonological loop; RAN = rapid automatized naming; syll. = syllable. All loadings from the latent constructs (PA, PL, RAN) to the corresponding indicators were significant.

The residual covariance between object naming and color naming (dashed line) was included in the reference group's model only.

[a] Factor correlation is non-significant at $p < .05$.