Chapter 2
Phonological Processing and Word Reading in Typically Developing and Reading Disabled Children: Severity Matters

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

This chapter is based on:
2.1 Introduction

Reading research of the past few decades has indicated phonemic awareness (PA) and rapid automatized naming (RAN) as strong predictors of word reading skill (Kirby et al., 2008; Landerl et al., 2013; Wagner et al., 2013). An influential theoretical framework involving both processes is Wolf and Bowers’ (1999) Double Deficit Hypothesis (DDH), which predicts that RAN and PA constitute more or less independent correlates of word reading ability. Since the introduction of the DDH, this assumption has been affirmed on numerous occasions (e.g., Compton et al., 2001; Kirby et al., 2010; Papadopoulos et al., 2009; Schatschneider et al., 2002; Torppa et al., 2012; Torppa et al., 2013; Vukovic & Siegel, 2006). The DDH also predicts that a combination of deficient RAN and PA – a “double deficit” – is associated with the poorest level of word reading performance (Wolf & Bowers, 1999, p.424). However, as DDH studies have employed differing criteria for RD, and have applied differing cutoffs to RAN and PA to indicate deficit levels, it remains unclear what should be considered as “the poorest level of word reading”, as well as what is meant with a double deficit. For instance, many DDH studies have focused on dichotomous RAN and PA deficit groups, whereas no reading ability levels are distinguished (e.g., Papadopoulos et al., 2009; Torppa et al., 2012; Torppa et al., 2013).

Similarly, most cross-sectional and longitudinal investigations of the predictive values of RAN and PA for reading performance (Kirby et al., 2003; Vaessen & Blomert, 2012; Van den Bos & De Groot, 2012) have also regarded reading performance across-the-board. Furthermore, the latter studies have suggested that the dynamics of RAN and PA change with age, where alphanumeric RAN typically becomes the dominant predictor over PA, while the influence of the latter declines with age. However, as no (fine-grained) distinctions are made for reading ability, these reports of developmental reading-predictive trajectories typically apply to the general reading population. Importantly, there are some developmental studies that did include dichotomous reader group comparisons (De Jong & Van der Leij, 2003; Van den Bos, 2008) which seem to modify the typical pattern of RAN becoming the dominant predictor over PA in two ways. First, RD children experience persistent RAN difficulties throughout primary education. Second, also PA remains disturbed in the RD groups, especially for more demanding tasks. These results suggest that compared to typically developing children, RD children show different reading-predictive patterns of RAN and PA that are relatively independent of age, and rather depend on reading ability level itself. Still, as these studies defined RD groups according to a single cutoff criterion,
further differential links of RAN and PA for multiple levels of reading ability possibly remain obscured.

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

Study aims and hypotheses

The guiding question that emanates from the rather diffuse and restricted insight into the RAN/PA-reading links mentioned above, is whether a stratified definition of reading (dis)ability would add to the general evidence as to how PA and RAN are related to word reading. The present study addresses this question with a systematic evaluation of these relationships for different ability levels along the whole word reading fluency continuum for a large sample of Dutch school children. Its main objective, therefore, is to investigate the reading-predictive patterns of PA and RAN as a function of word reading fluency while controlling for age.

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

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

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

2.2 Method

2.2.1 Participants
This study concerns a total of 1,598 Dutch children (813 ♀, 785 ♂) aged 7 to 14 years, mostly from the northern region of the Netherlands. The majority of the participants were recruited from intact classes of thirteen schools for regular primary and secondary education. Additionally, the sample contains a relatively small number of children with a formal diagnosis of Dyslexia, and/or either one of two other clinical conditions: Specific Language Impairment (SLI) and Attention Deficit Hyperactivity Disorder (ADHD). ADHD (American Psychiatric Association, 2000) and SLI (Tomblin et al., 1997) are considered as two of the most prevalent disorders in children. Because of the relatedness to reading of ADHD (e.g., McGrath et al., 2011; Willcutt et al., 2005), and SLI (e.g., De Bree et al., 2010; Vandewalle, Boets, Ghesquiere, & Zink, 2012), it seemed reasonable to include individuals from these groups when making general inferences about cognitive reading-related processes. The proportions of the resulting special groups are: dyslexia-only [n = 44 (2.8%)], ADHD-only [n = 15 (0.9%)], dyslexia + ADHD [n = 17 (1.1%)], SLI-only [n = 40 (2.5%)], dyslexia + SLI [n = 11 (0.7%)]. The children with SLI were sampled from schools for special education, specializing in language difficulties. The remaining special cases attended regular schools and were added to the present sample on the basis of referral by school psychologists or external health professionals. The special case groups may be considered to consist of the most severely impaired
children and it seems plausible that the remainder of the sample contains some undiagnosed (milder) cases as well. Children with significant hearing loss, uncorrected vision problems, neurological disorders, or low general cognitive functioning (i.e., IQ below 80) were excluded from the study.

Data collection

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

Reader-group assignments

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

2.2.2 Measurements

Word reading (WR). This standardized index of word reading fluency is based on assessments of real word- and pseudoword reading ability (Van den Bos & Lutje Spelberg, 2010; Van den Bos et al., 1994). For the present sample it is approximately normally distributed, with a mean of 10, and a standard deviation of 3 (Wechsler scores). Please refer to §1.6.1 for a detailed description.

Rapid Automatized Naming (RAN). Alphanumeric RAN was assessed with RANletters and RANdigits (Van den Bos & Lutje Spelberg, 2010). Separate standardized scores were computed for the subtests, as well as a standardized index of alphanumeric RAN (RANan). For the present sample, all three measurements are approximately normally distributed, with a mean of 10, and a standard deviation of 3 (Wechsler scores). Please refer to §1.6.2 for a detailed description.

Phonemic Awareness (PA). PA was assessed with the subtests PAelision, and PAsubstitution (De Groot et al., 2014). Separate standardized scores were
computed for the subtests, as well as a standardized composite index of PA (PAcom). For the present sample, all three measurements are approximately normally distributed, with a mean of 50, and a standard deviation of 10 (T scores). Please refer to §1.6.3 for a detailed description.

2.2.3 Data preparation

To facilitate the analyses, all test scores were converted to z-scores. Also, to avoid confounding of the results for RAN and PA, principal component analysis (PCA) with varimax rotation and Kaiser normalization was performed on the RAN subtasks and the PA subtasks (see also Table 2.1). This resulted in two orthogonal components that clearly bear on RAN and PA, referred to as RANFAC and PAFAC, respectively. Eighty-five percent of total variance was explained by these components. However, as these orthogonal components do not address the commonality of RAN and PA, in addition, an encompassing composite variable, referred to as RANPA, was created by averaging RAN_{an} and PA_{com}. Thus, three variables have been analyzed, PAFAC, RANFAC and RANPA.

Table 2.1 Rotated factor loadings on RAN-factor (RANFAC) and PA-factor (PAFAC)

<table>
<thead>
<tr>
<th>Component</th>
<th>RANFAC</th>
<th>PAFAC</th>
</tr>
</thead>
<tbody>
<tr>
<td>RAN\text{digits}</td>
<td>.93</td>
<td>.14</td>
</tr>
<tr>
<td>RAN\text{letters}</td>
<td>.90</td>
<td>.24</td>
</tr>
<tr>
<td>PA\text{elision}</td>
<td>.26</td>
<td>.86</td>
</tr>
<tr>
<td>PA\text{substitution}</td>
<td>.12</td>
<td>.91</td>
</tr>
</tbody>
</table>

2.2.4 Statistical analyses

Three different statistical procedures were employed to evaluate the hypothesized reading-related effects for RANFAC, PAFAC, and RANPA. The procedures are: (1) computing standardized effect sizes (Cohen’s $d$), (2) conducting univariate and multivariate analysis of variance ((M)ANOVA), and (3) evaluating receiver operating characteristics (ROC).

First, Cohen’s $d$ reveals the magnitudes of main effects for RANFAC, PAFAC, and RANPA for each reading ability level. Values of .2, .5, and .8 are considered as small, medium, or large effect sizes, respectively (Cohen, 1988). These effect sizes will be used for the evaluation of the magnitudes of hypothesized group differences. However, these effect sizes do not provide information on variance components, i.e., the unique and common contributions to reading proficiency, and on (predicted) interaction effects. For this, analysis of variance was deemed appropriate.
Therefore, the second, and most central method of the present study is (M)ANOVA, which was used to evaluate main and interaction effects for each WR-dichotomy. Separate runs ($8 \times 3 = 24$) of the analyses were performed with the two orthogonal components RANFAC and PAFAC, and the standardized mean of PAc and RANan (RANPA), as dependent variables, and the WR dichotomies as independent variables. This method also allowed to control for age. As the standardized scores underlying the dependent variables are based on age levels corresponding to years or semesters, the models were corrected for residual age effects by including age in months as a covariate.

To take into account obvious inequalities in terms of the reader group sizes, resulting from the application of the WR criteria, a resampling procedure comprising 1000 iterations with the WR-dichotomy as the stratification variable was performed. This procedure checks each RD or ER group against 1000 reference groups (REF) of equal size and averages the effect sizes. To investigate the interaction effects between WR-dichotomy, PAFAC and RANFAC, a similar sampling procedure was applied to procure equal numbers of children from the RD and ER groups. This downsized all groups to $n = 68$, with the exception of the ‘$M + 2.0SD$’ ER group ($n = 30$, see also Appendix A) which was oversampled with a factor 2.27. Effect sizes for the (M)ANOVAs are presented as $\eta^2$, with values of .01, .06, and .14 considered to be small, medium, or large effect sizes, respectively (Cohen, 1973). Note that these analyses were repeated without the resampling procedure, yielding slightly attenuated effect sizes, relative to the reported values for the resampled data.

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

2.3.1 Descriptive analysis

In Figure 2.1 the standardized means (z scores) of the two phonological components are plotted against the WR conditions.

![Figure 2.1 Plotted means (z-scores) with error bars (SEM) of RANFAC and PAFAC for RD/ER versus REF groups by WR dichotomy (-2.0sd through +2.0sd) at the horizontal axis](image)

It is evident that the means of RAN factor (RANFAC) and the PA factor (PAFAC) show distinct patterns for the target groups across the different WR criteria. Judging from the slopes, it appears that both measures interact with the WR criteria, as their means deviate from the population mean more strongly towards the more extreme criteria. Finally, the apparent widening pattern towards the more extreme WR criteria for the RD side of the continuum (WR-SDRD) is suggestive of a possible three-way factor by condition interaction, i.e., $RANFAC \times PAFAC \times WR-SD_{RD}$. We will return to this issue in the next section.

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

Regarding the group differences in terms of mean differences, the effect sizes (Cohen’s $d$) of group membership on RANFAC, PAFAc, and the composite variable RANPA are plotted against the WR criteria on the horizontal axis in Figure 2.2.

From Figure 2.2 it is evident that overall effect sizes are at least substantial. There is a clear enduring influence of RAN and PA throughout the reading fluency continuum. Whether an effect is considered as positive or negative is indicated by the sign of the WR criterion (see also Figure 2.1). Second, RANPA yields substantial added value over the separate components in terms of predictability. Moreover, RANPA clearly is more affected in the RD groups, which seems mostly due to elevated effect sizes for PA, as opposed to a rather stable pattern for RAN across most conditions. Still, effect sizes for both components become markedly larger towards the more negative conditions, which is suggestive of an interaction with WR-SD RD. Therefore, these results may be taken as descriptive evidence in support of Hypothesis 1.

Regarding the separate components, PAFAc seems more affected in the most severe RD group, relative to RANFAC, lending support to Hypothesis 2. Finally, perhaps the most intriguing observation about Figure 2.2 is the ostensive drop in impact for PAFAc when going from the negatively to the positively oriented WR criteria. Apparently, the relative dominance of PAFAc

Figure 2.2  Standardized effect sizes of mean differences (Cohen’s $d$) with error bars (SEM) between RD/ER and REF groups for RANFAC, PAFAc, and RANPA by WR dichotomy
over RANFAC for RD children, reverses into a dominance of RANFAC over PAFAC for ER children. This may be taken as descriptive evidence for the Hypothesis 3.

2.3.3 Analysis of variance

Table 2.2 summarizes the results for the MANOVAs that were performed on all eight WR dichotomies (WR-SDALL). Generally, these results indicate strong group predictability for both RAN and PA throughout the WR continuum, as is reflected by overall large proportions of explained variance ($\eta^2$) for each dependent variable (see A, B, & C in Table 2.2). Additionally, there is evidence of under-additivity in terms of marginal variance explained by RANPA (see C in Table 2.2), as the cumulative effects of the orthogonal components RANFAC and PAFAC – particularly those of the lower bound – clearly outweigh the combined effects of RANan and PAcom (RANPA). Furthermore, these results indicate a quite stable pattern across all eight dichotomies for large proportions of explained variance pertaining to RANFAC, albeit somewhat more for the strictest RD criteria. In contrast, the stability found for RANFAC clearly does not apply to PAFAC. PAFAC yields large effects for RD with the largest effects found for the most strict RD criteria, whereas markedly smaller effects were found for the above-average readers.

To evaluate the interaction effects between WR-SDALL, RANFAC and PAFAC, three separate ANOVAs were performed on each dependent variable, with WR criterion as the grouping factor, covering (i) four levels for RD (WR-SDRD), (ii) four levels for ER (WR-SED), and (iii) all eight levels of the (whole) WR continuum (WR-SDALL). Table 2.3 summarizes the results, presenting $F$ ratios, $p$ values, and effect sizes ($\eta^2$) for each predictor, according to the range of included WR criteria.

<table>
<thead>
<tr>
<th>WR-SDRD</th>
<th>WR-SED</th>
<th>WR-SDALL</th>
</tr>
</thead>
<tbody>
<tr>
<td>F</td>
<td>p</td>
<td>$\eta^2$</td>
</tr>
<tr>
<td>RANFAC</td>
<td>3.56</td>
<td>.015</td>
</tr>
<tr>
<td>PAFAC</td>
<td>6.35</td>
<td>&lt;.001</td>
</tr>
</tbody>
</table>

Note: degrees of freedom for WR-SDRD and WR-SED are: $F_{(3, 269)}$, and for WR-SDALL: $F_{(7, 536)}$. 

38 Chapter 2
Table 2.2  F ratios and effect sizes (η²) for RANFA, PAFAC, and RANPA variance components per WR dichotomy (WR-SD).

<table>
<thead>
<tr>
<th></th>
<th>RANFA (A)</th>
<th>PAFAC (B)</th>
<th>PAFAC (C)</th>
</tr>
</thead>
<tbody>
<tr>
<td>+4.3</td>
<td>1.3</td>
<td>1.5</td>
<td>2.0</td>
</tr>
<tr>
<td>+1.7</td>
<td>1.3</td>
<td>1.5</td>
<td>2.0</td>
</tr>
<tr>
<td>+1.5</td>
<td>1.3</td>
<td>1.5</td>
<td>2.0</td>
</tr>
<tr>
<td>+1.5</td>
<td>1.3</td>
<td>1.5</td>
<td>2.0</td>
</tr>
<tr>
<td>+1.5</td>
<td>1.3</td>
<td>1.5</td>
<td>2.0</td>
</tr>
<tr>
<td>+1.3</td>
<td>1.3</td>
<td>1.5</td>
<td>2.0</td>
</tr>
<tr>
<td>+1.3</td>
<td>1.3</td>
<td>1.5</td>
<td>2.0</td>
</tr>
<tr>
<td>+1.3</td>
<td>1.3</td>
<td>1.5</td>
<td>2.0</td>
</tr>
<tr>
<td>+1.3</td>
<td>1.3</td>
<td>1.5</td>
<td>2.0</td>
</tr>
<tr>
<td>+1.3</td>
<td>1.3</td>
<td>1.5</td>
<td>2.0</td>
</tr>
</tbody>
</table>

Note: All F-ratios significant with one degree of freedom and p < .001.
Regarding the set of ANOVAs for the RD cutoffs (i), a significant three-way \textit{RANFAC} x \textit{PAFAC} x \textit{WR-SD} \textit{RD} interaction, $F_{(3, 268)} = 2.49, p = .040$, was found, indicating differential patterns for \textit{RANFAC} and \textit{PAFAC} as a function of \textit{WR-SD} \textit{RD}. Decomposing this effect into two separate two-way interactions, a moderately strong \textit{RANFAC} x \textit{WR-SD} \textit{RD} interaction ($\eta^2 = .04$) and an even more substantial \textit{PAFAC} x \textit{WR-SD} \textit{RD} ($\eta^2 = .07$) interaction effect were found (see Table 2.3). This difference is in line with the results presented in Table 2.2, indicating that group membership explains accrued proportions of \textit{PAFAC} variance for the strictest criteria. Indeed, as compared to the less strict \textit{WR-SD} (M – 1.5SD) and \textit{WR-SD} (M – 1.3SD) criteria, post hoc analysis with Tukey HSD correction indicated significant \textit{PAFAC} differences for \textit{WR-SD} (M – 1.7SD) (p = .056; p = .003, respectively), and \textit{WR-SD} (M – 2.0SD) (p = .054; p = .003, respectively) only. Regarding \textit{RANFAC}, a similar, but less strong pattern applies, with only the contrasts with \textit{WR-SD} (M – 1.3SD) yielding (marginally) significant differences (\textit{WR-SD} (M – 1.7SD), p = .06; \textit{WR-SD} (M – 2.0SD), p = .006). In conclusion, the finding of substantial \textit{PAFAC} x \textit{WR-SD} \textit{RD} and \textit{RANFAC} x \textit{WR-SD} \textit{RD} interaction effects can be interpreted as support for Hypothesis 1. The combination of a significant three-way interaction, and, relative to \textit{RANFAC} x \textit{WR-SD} \textit{RD}, a large \textit{PAFAC} x \textit{WR-SD} \textit{RD} interaction effect can be interpreted as support for Hypothesis 2.

With regard to the set of ANOVAs for the ER cutoffs (ii), and in contrast to the previous results, \textit{PAFAC} clearly is less strongly (positively) affected for the above-average readers. It can be noted, however, that there is a moderately strong \textit{PAFAC} x \textit{WR-SD} \textit{ER} interaction effect ($\eta^2 = .04$) (see also Table 2.3), which can be interpreted as a small increase of effect size for \textit{PAFAC} towards the more extreme \textit{WR-SD} \textit{ER} criteria. The interaction with \textit{RANFAC} was not significant. In line with Hypothesis 3, the finding of persisting effects for \textit{RANFAC} along with generally diminished effect sizes for \textit{PAFAC}, suggests that word reading is relatively more closely associated with \textit{RAN} than with \textit{PA} in the ER groups. Additionally, it can be noted that, whereas the common variance components are relatively small, there is a clear difference between the unique variance components of \textit{RANFAC} and \textit{PAFAC}, with \textit{RANFAC} values being considerably larger (see bottom half of Table 2.2).

Testing the interactions of \textit{RANFAC} and \textit{PAFAC} for the whole continuum (i.e., \textit{RANFAC} x \textit{WR-SD} \textit{ALL} and \textit{PAFAC} x \textit{WR-SD} \textit{ALL}) (iii), yielded substantial effects: $\eta^2 = .20$, and $\eta^2 = .18$, respectively (see Table 2.3). Together with the pattern of increasing proportions of unique variance pertaining to \textit{RANFAC} and diminished proportions for \textit{PAFAC} when going from \textit{WR-SD} \textit{RD} to \textit{WR-SD} \textit{ER} (see Table 2.2), these results can be interpreted as strong support for Hypothesis 3.
2.3.4 ROC analyses

The previously mentioned magnitude differences and interaction effects are also made apparent by the graphical sequence depicted in Figure 2.3 (next page), where the ROC curves of RANFAC, PAFAC, and RANPA are plotted across the WR cutoffs. The reference line (diagonal) represents pure chance, and a measurement producing a curve near or even below the diagonal can be thought of as chance performance or even worse. Conversely, the higher the area under the curve (AUC) and the steeper the slope – implying increased sensitivity at little expense to the specificity – the more effective a measure is in predicting group membership. Demonstrated by decreasing AUCs, Figure 2.3 clearly shows a pattern of waning PAFAC predictability as word reading skill increases, whereas RANFAC yields a much more stable pattern. More generally, going from negatively to positively oriented dichotomies, the reversal in terms of the components’ predictive values is obvious. These results may be taken as additional support for Hypothesis 3.

This pattern is also apparent from the high AUCs (see Table 2.4), sensitivities, and specificities. The highest AUC value is found for RANPA (.96) at the strictest RD criterion (WR-SDM - 2.0SD). If one opts for a 95% accurate reader group prediction (accuracy = true positives + true negatives / total n) (cf., Fawcett, 2006), the optimal cutoff score for RANPA is M - 1.5SD. This cutoff yields a sensitivity of .77 and a specificity of .95. As a final remark, it can be noted that, for this criterion the PAFAC curve approximates the RANPA curve, indicating a stronger impact for PAFAC relative to RANFAC. Together, these results may be taken as additional support for Hypothesis 2.

<table>
<thead>
<tr>
<th>Table 2.4 Areas under the curve (AUCs) for RANFAC, PAFAC, and RANPA per WR dichotomy (WR-SD)</th>
</tr>
</thead>
<tbody>
<tr>
<td>WR-SD</td>
</tr>
<tr>
<td>RANFAC</td>
</tr>
<tr>
<td>PAFAC</td>
</tr>
<tr>
<td>RANPA</td>
</tr>
</tbody>
</table>
Figure 2.3  ROC curves for RANFAC, PAFAC, and RANPA by WR-criterion
2.4 Discussion

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

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

The second motive for the present study and the adopted approach of differentiating for reading ability levels, was to include the virtually forgotten group of highly proficient readers. In terms of theoretical underpinnings as well as treatment goals of RD, it seems important to ask what can be learned from those who clearly have mastered decoding. In line with general reading theory (De Jong & Van der Leij, 2003; Ehri, 2005; Gough, 1996; Van den Bos, 2008) it was hypothesized that highly proficient reading, as opposed to deficient reading, is characterized by a relatively strong link to highly automatized access to alphanumeric networks (as measured by RAN\textsubscript{an}), and a decreased necessity of sub-lexical analytical processes (as measured by PA). We have found strong evidence to support this hypothesis, as our data indicate a striking reversal in
terms of RAN and PA deviation means (Figure 2.1), as well as effect sizes (Figures 2.2 and 2.3), when switching from reading-disabled to above-average reader groups. It should be noted, however, that the dominance of RAN over PA in above-average reader groups is relative and due to diminished effect sizes of sub-lexical processing in the presence of a rather stable contribution of RAN across word reading ability levels. Thus, in above-average readers, RAN seems to be the dominant predictor by default. The finding of RAN being a rather stable correlate of reading fluency across ability levels, is a new one, at least in comparison to previous non-developmental, and developmental studies for the general population. For example, in their meta-study Swanson, Trainin, Necoechea, and Hammill (2003) reported that RAN-Reading correlations varied only minimally with age, but that these correlations were weaker in poor readers than in skilled readers. Other studies, however, report increasing correlations with age (Vaessen & Blomert, 2010; Van den Bos & De Groot, 2012). Interestingly, our results neither indicate an increase with age, nor a decrease with impairment. For future studies, it would, therefore, be quite informative to cross validate the present findings by means of a comprehensive developmental design that discerns reading ability levels as well.

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