Psychology Science Quarterly, Volume 51, 2009 (3), pp. 315-338

Does systematic reading instruction impede prediction of reading a shallow orthography?

ALFRED SCHABMANN[1], BARBARA MARIA SCHMIDT[1], CHRISTIAN KLICPERA[1], BARBARA GASTEIGER-KLICPERA[2] & KATHRIN KLINGEBIEL[3]

Abstract

Different methods of reading instructions have been the subject of controversy. This study examined the influence of systematic phonics vs. non-systematic phonics methods of instruction on the prediction of reading. 443 kindergarten children were tested on phonological awareness, naming speed and visual word matching using the Bielefelder Screening (BISC). Children were retested in grades one and four. Results showed that although the prognostic validity of kindergarten measures was generally low, it was considerably higher for grade one children in classes with non-systematic phonics instruction. Children who received systematic phonics instruction scored significantly higher on measures of phonological decoding as compared to their peers who received less systematic instruction. Implications for the prediction of reading and early screenings are discussed.

Key words: reading instruction; prediction of reading; phonics instruction; phonological awareness; rapid automatized naming

[1] Correspondence concerning this article should be addressed to: ao. Prof. Dr. Mag. A. Schabmann, Fakultät für Psychologie, Institut für Wirtschaftspsychologie, Bildungspsychologie und Evaluation, Universitätsstraße 7, Top B0606, 1010 Wien, Austria; email: alfred.schabmann@univie.ac.at

[2] University of Education, Weingarten, Germany

[3] University of Sussex, Brighton, United Kingdom

Introduction

Early identification of children at risk for reading difficulties is of particular importance. It allows timely intervention to avoid serious reading problems which could otherwise cause failure during early schooling and later academic career (e.g. Schabmann & Kabicher, 2007; Torgesen & Burgess, 1998).

Previous research in English speaking countries has demonstrated that reading achievement can be predicted by early measures of phonological awareness (PA) and naming speed. PA implies the ability to manipulate large (e.g. syllables) and small (e.g. phonemes) units of spoken language, which helps children to understand the grapheme-phoneme coding principle in written language (e.g. Bradley & Bryant, 1985; Bryant, Bradley, MacLean & Crossland, 1989). Early difficulties in PA hence interfere with the acquisition of alphabetic coding, particularly at the beginning of learning how to read. PA tasks include rhyming, syllable counting, phoneme deletion, phoneme blending, and phoneme detection (for further tasks see Landerl & Wimmer, 2000; 2008).

To measure naming speed subjects are asked to recall and name letters, digits, objects, or colours as quickly as possible (also called rapid automatized naming, RAN). So far, no consensus about the nature of relationship between RAN and reading has been found. Some authors argue RAN should be considered part of a more general "phonological processing" construct that reflects the access to phonological information in the long term memory (e.g. Torgesen, Wagner & Rashotte, 1998). However, previous studies showed that RAN accounts for additional variance in reading even when controlling for phonological awareness (Bishop, 2003; Blachman, 1984; Bowers, 1995; Bowers & Swanson, 1991; Compton, DeFries & Olson, 2001; Cornwall, 1992; Landerl & Wimmer, 2008; Manis, Doi & Badha, 2000; McBride-Chang & Manis, 1996; Wimmer, 1993; Wolf & Bowers, 1999).

Some authors consider RAN as an indicator of a mechanism that is essential for automatic processing of orthographic (rather than phonological) representations in memory, i.e. slow letter identification would make it difficult for children to acquire the sensitivity for frequently occurring orthographic patterns (e.g. Bowers & Wolf, 1993; Wolf, Bowers & Biddle, 2000; for a more detailed discussion see Kirby, Parilla & Pfeiffer, 2003 or Georgiou, Parilla, Kirby & Stephenson, 2008; for a recent finding challenging this view see Moll, Fussenegger, Willburger & Landerl, 2009). Other authors (e.g. Spring & Davis, 1988) assume that RAN might reflect the ability to coordinate multiple concurrent processes (i.e. visual, linguistic and articulatory) or general cognitive processing speed (e.g. Kail, Hall & Caskey, 1999).

For the predictive strength of early PA and RAN measures at least three methodological aspects are critical (Verhagen, Aarnoutse & van Leeuwe, 2008): The manner in which PA and RAN are measured, the implemented reading measures (i.e. reading accuracy vs. reading speed), and the regularity of the orthographic system (i.e. the transparency of letter to sound mapping, see Landerl & Wimmer, 2008). In the present study, we primarily concentrate on the latter. Languages differ in terms of orthographic regularity. English, for example, contains many inconsistencies in grapheme-phoneme mapping and is therefore regarded as *deep* orthography. On the other hand, German with its very consistent grapheme–phoneme correspondences is regarded as *shallow* orthography. The current findings are summarized below:

1. In languages with relatively less consistent orthography the effect of PA on the later reading accuracy and reading speed seems stronger and most notably more stable than in

consistent orthographies. For instance, while both PA and RAN predict reading accuracy in English-speaking countries until grade three (e.g. Parilla, Kirby & McQuarrie, 2004; Schatschneider, Francis, Carlson, Fletcher & Foorman, 2004), PA was only predictive for grade one reading accuracy in a more regular orthography (Dutch) in the study of Verhagen, Aarnoutse & van Leeuwe (2006). Landerl and Wimmer (2008) showed that PA predicted reading fluency in Austrian children in grade one, but failed as a predictor in grade four and eight. However, contradictory findings were reported in two cross-cultural studies (Caravolas, Violín & Hulme, 2005; Patel, Snowling & de Jong, 2004) in which PA contributed to the prediction of reading speed and accuracy for English-speaking children as well as children speaking a more regular language in terms of orthographic consistency (Dutch and Czech).

2. For English-speaking countries (or countries with less consistent orthography) findings on the prognostic strength of RAN measures for reading speed were inconsistent (see Verhagen et al., 2008). Yet, RAN was frequently found to be a strong predictor of reading accuracy, even though it appeared to contribute less to the prediction of reading when objects were used as stimuli rather than letters or digits (e.g. Manis et al., 2000; Neuhaus & Swank, 2002). For more regular orthographies, RAN was found a stable predictor for reading accuracy and reading speed (e.g. Landerl & Wimmer, 2008; Verhagen et al., 2008; Wimmer, Mayringer & Landerl, 2000).

Nevertheless other aspects may be of additional importance (see e.g. Verhagen et al., 2008). The limited predictive strength of PA in regular orthographies with consistent effects of RAN can be interpreted as the outcome of two interacting components of learning to read in these language systems. Firstly, as grapheme-phoneme correspondences (GPC) are very transparent and hence leave little room for ambiguity, it is relatively easy for young German readers to map graphemes to phonemes properly, even for children with low PA preliminary to reading instruction. One year after the onset of instruction, German-speaking children make on average less than 5% reading errors and even the weakest readers in grade two with a percentile <5 read eight out of ten words correctly (Klicpera & Schabmann, 1993). In other words, a deeper insight of the phonological structure of spoken language is not necessary prior to reading instruction to acquire the basic principles of alphabetic reading, especially if reading instruction focuses on systematic introduction of GPC. Secondly, recent findings suggest that PA could be rather a consequence of early reading instruction than a precondition for learning to read, and that children acquire PA simultaneously with alphabetic reading (e.g. Cossu et al., 1988; Morais, Alegria & Content, 1987; Wimmer et al., 1991). This is particularly the case when PA is measured on phoneme level. Wimmer et al. (1991) found that although almost all of their participating children failed in a vowel substitution task before reading instruction, most of them scored close to perfect in this task after only a few months of school. Furthermore, it was found that children with good PA prior to reading instruction exhibited high reading accuracy in school, however, the same was true for nearly all the children with little or no PA at the beginning of grade one. The authors conclude that "phonemic awareness should not be considered a precondition which must be met before successful reading and spelling instruction can begin" (Wimmer et al, 1991; p. 245). For them the critical point is how easy or difficult PA can be induced by reading instruction.

If this assumption is true, reading instruction itself becomes an important aspect in the early prediction of reading. In systematic letter-to-sound instructions children with little or no PA prior grade one would still have the chance to acquire both PA and alphabetic reading solely driven by the received reading instruction (although a few might still fail). On the other hand, if a less systematic instruction is given, children with little or no PA prior grade one would fail to develop PA and hence alphabetic reading; although they have potential to be successful in both. Thus, in classes with less systematic letter-to-sound instruction the correlation between early PA and reading will be stronger than in classes with a more systematic letter-to-sound instruction. This argument was brought by Ehri (1993). “It would seem that with carefully planned instruction and sufficient practice to insure the learners master the alphabetic system and become able to use in automatically to decode new words and to acquire sight words, the predictive success of measures like phonological awareness […] might be weakened because good instruction should remediate deficits and hence eliminate the influence of individual differences on the literacy acquisition process.” (Ehri, 1993, p. 249).

As described above, the cognitive mechanisms underlying the RAN-reading relationship are not entirely understood. Two competing conclusions can be drawn from the theoretical positions previously cited regarding to the influence of reading instruction on the predictive strength of preschool RAN measures: If RAN is conceptualized as an indicator of the cognitive processing (which might not be touched by reading instruction) it would be a predictor of reading *irrespective of the instructional method*. On the other hand, if the assumptions of the double-deficit theory (e.g. Wolf, Bowers & Biddle, 2000) are correct and RAN reflects orthographic processing, a systematic letter-to-sound instruction should be helpful to develop an orthographic sensitivity for frequently occurring letter strings and thus would diminish the predictive strength of RAN, predominately at the later stages of reading acquisition and for reading speed.

Research questions

Although the (possible) importance of systematic reading instruction is often stated (e.g. Landerl & Wimmer, 2008; Marx & Weber, 2006; Wimmer et al., 1991) there are little findings about this issue in German readers. A few studies were carried out in the late 60s (Ferdinand, 1970; Holzinger, 1964; H. Müller, 1964; R. Müller, 1965; Schubenz, 1966): Classes with systematic phonics instruction were compared to classes in which whole-word (i.e. non phonics) instruction methods were the primary approach. Differences in reading and spelling scores showed that children who received systematic phonics instruction outperformed their peers in classes with whole-word instruction. However, differences did not persist until the end of grade four. Furthermore, in classes with whole-word instruction, progress in reading and spelling seemed to depend more on the general intellectual ability (Holzinger, 1964; H. Müller, 1964; Schubenz, 1966).

One reason for the lack of studies might be the common use of a letter-to-sound approach in German speaking countries. Whole-word reading instruction was abandoned almost completely in the late seventies. Yet, systematic phonic instruction shows great variation in how it is implemented in class. For instance, within the rather broad framework of the curriculum for primary schools in Austria, teachers are free in their choice of instruction method, i.e. they can decide which primers or other materials to use. They are even allowed to refuse the use of primers at all while using their own materials (e.g. lists of words from the internet or other sources). Taking this information into account it seems feasible to assume that there are classes in which the letter-to-sound instruction is considerably less systematic than in others.

In the present study we explored the influence of instruction on the prediction of reading achievement by comparing extreme groups based on which materials (e.g. primers, sight word lists) teachers used and how they proceeded with instruction (see below for a more detailed description). Schabmann (2007) showed that the use of these criteria is a feasible and theoretically appropriate approach for identifying classes in which sufficiently different teaching methods in terms of letter-to-sound teaching (i.e. explicit teaching of the alphabetic principle) are employed. We identified classes with highly systematic phonics instruction and compared them to classes with less systematic instruction in grade one. In the following we will refer to them as SP (systematic phonics instruction) and NSP (less/non systematic phonics instruction).

In the interest of practical relevance we used a well known and widely accepted method for prediction, *The Bielefeld Screening Battery to predict reading and spelling difficulties* (Bielefelder Screening zur Früherkennung von Lese- Rechtschreibschwierigkeiten: BISC; Jansen, Mannhaupt, Marx & Skowronek, 1999). This screening uses four PA tasks and two RAN tasks, which are described below. In addition a visual word matching task and a phonological decoding task (repetition of nonwords) are presented, as the authors propose that the processing of text requires controlled visual attention (word matching) as well as the recall of phoneme sequences previously stored in the working memory (phonological decoding). According to the authors, the screening battery proved highly effective in predicting students at risk for later reading and spelling difficulties. However, recent studies show contrary results. For instance, Marx and Weber (2006) compared reading, spelling, reading comprehension and mathematic test scores of grade one to grade four students, either classified "at risk" or "not at risk" according to the BISC four months prior to school entry. The mean effect size for reading achievement (reading fluency) was d=.62, yet the percentage of children who had reading problems at school and were identified correctly "at risk" in kindergarten ranged from only 32% (grade one) to 47% (grade two). Effectively only 19% (grade two and three) to 25% (grade four) of the children "at risk" had reading problems in school.

In the following we will report our results on the prognostic validity of the BISC. In addition we will analyze how its main components (i.e. PA, RAN, visual word matching task) succeed in predicting later reading achievement and to what extent the instructional method is of influence. We presented lists of words with different frequencies in written language, and lists of nonwords. Our assumption was that reading instruction will have different effects on target words which differ in the frequency of their occurrence in written language, particularly when PA is used as a predictor. When nonwords and low frequent words are used as reading measures the demands on phonological decoding skills will be high and the method by which letter-to-sound mapping is taught will become important. On the other hand, when very frequent and familiar words are used, the effect of instructional variations may be low, because children in NSP may rely stronger on global visual characteristics of the target word (measured by the visual word matching task), as this strategy is forced by non-systematic reading instruction alternatively to phonological decoding.

In summary, we expect a stronger prediction for NSP than for SP in reading accuracy shortly after the beginning of reading instruction. This effect should be more distinct for PA than for other predictors, for PA is thought to have the most direct influence on phonological decoding which in turn is directly influenced by reading instruction. No detailed hypothesis as to how long this difference may persist is stated, but it is believed that NSP will catch up in their phonological skills and therefore by grade four reading scores should correlate with preschool PA measures essentially similar for both SP and NSP classes. There are two reasons for the latter assumption. Firstly, due to the relatively high consistency of the German language even in a NSP instruction most of the children who show little or no preschool PA are expected to catch up. If so, a self teaching mechanism sensu Jorm and Share (1983) could be induced as a "starting point" for the independent acquisition of an autonomous orthographic lexicon and therefore skilled (and fast) word recognition. Secondly, despite all differences, NSP instruction is still a form of *phonics instruction*, i.e. teachers might sooner or later introduce GPC, although this might take longer and happen in a far less systematic manner compared to SP (for a more detailed description of NSP and SP classes see below). Therefore, only a delay in reading acquisition is expected, not a persistent impairment for students in the NSP group.

In line with previous findings, we expect RAN to be a stronger and more persistent predictor for *reading speed* than PA. Due to technical reasons we were not able to measure reading speed at the beginning of grade one, so it was not possible to make predictions about differences between SP and NSP at this early stage. However, we were able to consider differences at the end of grade one and even grade four. If decoding skills induce a self-teaching mechanism that will lead to skilled and automatic reading, children in NSP classes should be delayed in reading speed as they are expected to develop decoding skills later than their peers in SP classes.

Method

Sample

443 children (220 boys) were tested with the BISC in kindergarten four months before school. 421 were retested on their reading skills again after three months of reading instruction (in November). All 421 children were tested again at the end of grade one, and 370 (190 boys) of these children were followed up until the end of grade four. Intelligence was tested in grade three. Children were excluded if they had to repeat a grade, if German was not their first language or if they were identified as intellectually disabled by the school authority. All children came from small towns or townships in Lower Austria. They were comparable in terms of socio-economic background, and were from similar regional environments. On average, children were 6.4 years old at the time of the first testing.

Procedure

Classification of SP and NSP: The following criteria were used to assess the extent to which a systematic letter-to-sound instruction was given to the children.

To be classified as SP, teachers had to

1. Use a primer in which letter-to-sound mapping was introduced systematically: Letters were formally introduced prior to word reading. Words which appeared in the primers were solely composed of these letters.

2. Teachers worked through the primer lesson by lesson in order of appearance: This criterion was found useful in earlier studies (e.g. Schabmann, 2007).
3. Children were not urged to read words which contained unknown letters until all letters were introduced (In fact in some primers words which contained unknown letters occurred. However, they were marked by colours or pictures next to them to indicate that these words should be named by the teacher without the children being forced to decode them).

For NSP the criteria were:

1. Teachers used a primer in which letter-to-sound mapping was not introduced systematically: In this group, some letters were introduced prior to reading words, but children were also urged to read words which contained unknown letters.
2. Alternatively: Teachers used individually assembled lists of sight words. Children were required to individually learn GPC by repeatedly reading similar words.
3. There was no identifiable order of letter instruction. Many letters were introduced "on demand", i.e. when teachers became aware of decoding problems.

Out of the 49 classes 9 were identified as NSP and 37 as SP. Three classes (46 children) could not be assigned, because the method of reading instruction did not match one or the other criteria, primarily because the teachers did not work through SP lessons in order of their appearance in the book or single lessons where skipped. In NSP classes 99 children (44 boys) remained with complete data for comparison. In SP classes reading was tested on 276 children (150 boys). Children in NSP and SP classes did not significantly differ in age (NSP: mean = 6.5 years, std. = .49; SP: mean = 6.4 years, std. = .36). At the end of grade four 91 (41 boys) children in NSP and 242 (131 boys) children in SP were retested.

Measures

Cognitive ability and pre-reading skills: To control for non-verbal cognitive ability, intelligence was measured using the Standard Progressive Matrices Test (SPM; Kratzmeier & Horn, 1988) in third grade.

Predictors of reading: Predictors of reading were tested in kindergarten using the Bielefelder Screening (BISC; Jansen et al., 1999) four months prior to the beginning of grade one. The screening includes four different types of tasks:

1. *Phonological awareness:* Two subtests assess large grain-sized phonological awareness (syllables and rhymes) and two subtests phonological awareness on phoneme level (phoneme blending and phoneme analysis). *Rhyming task:* Children have to decide whether two spoken words rhyme. *Syllable segmentation task:* Children are asked to repeat words and clap the rhythm. *Phoneme blending:* A word is presented by isolating a relevant phoneme (e.g. Z-ange [English: pliers]). On a response card 4 images of objects are shown and the child has to point to the object that corresponds with the presented word. *Phoneme analysis:* The child is supposed to indicate whether a certain phoneme is part of a presented word. For each task, the total of correct answers is scored.
2. *Naming Speed:* The children are supposed to name the appropriate colours for a total of four recurrent and familiar objects (types of fruit) that are presented as black and white

images. In a second task which is supposed to verify the liability to interference of the rapid naming process (see Marx, 1985), the same objects are presented as inappropriately coloured images (e.g. blue lemon). The time used to complete the two tasks is scored.

3. *Phonological decoding in short-term memory:* Children are asked to repeat 3-5 syllable nonwords (e.g. RISOLAMU). The total of correct answers is scored.
4. *Visual word identification:* On one card a target word is presented with four alternative words out of which one is identical to the target word. This word has to be identified. The number of correct responses as well as speed is measured.

For the calculation of “risk points”, all subtests were used and children “at risk” were identified according to the instruction in the BISC manual, “borderline cases” (three “risk points”) were assigned to the risk-group. For further examination an exploratory factor analysis (PCA, Varimax) was carried out on the subtests of the BISC. Three factors were extracted that accounted for 54% of the variance: phonological awareness (PA), visual attention (VIS), and rapid naming (RAN). Contrary to our expectations and the theoretical structure of the screening, the syllable task appeared to load only weakly on the PA factor. Further analysis of this task revealed a low reliability, hence we decided to skip this task and run a second analysis in order to receive adequate factor scores for the three factors to use for regression analysis. Loadings and communalities are given in table 1.

November grade one reading accuracy: For early reading, a newly developed reading test was used (see Klicpera, Humer, Gasteiger-Klicpera & Schabmann, 2008). It consists of lists of 28 words and 16 nonwords. The words were presented in blocks of 4 words, each on a sheet of white paper A4 in AvantGarde Bk BT (black, size 72pt), as this font is very similar to the one used in most primers in Austria. The word length varies from 2 to 5 letters for both words and nonwords. Of the 28 words 16 had already been introduced in textbooks or sight-word lists at the time of testing. 12 words were unfamiliar to the children, i.e. they did

Table 1:
Communalities, percentage of explained variance and factor loadings for the subtests of the BISC which were used for further statistical analysis

	h^2	PA	RAN	VIS
Percentage of variance		25.3	15.6	13.6
nonword repetition	.54	.73	.01	-.04
phoneme blending	.48	.66	-.06	-.08
rhyming	.55	.58	.23	.39
phoneme analysis	.40	.47	.21	.38
visual word identification accuracy	.66	.01	.85	-.15
visual word identification speed	.74	.07	.78	.22
naming speed black and white	.47	-.14	-.07	.72
naming speed coloured	.52	.15	.08	.67

not appear in the primers or word lists. All words and nonwords were composed of the first 8 known letters. SP teachers were asked to mark the lesson in the primer which they had reached 2 weeks prior to testing, so it could easily be identified which letters were already known to the children. If the children already knew more than 8 letters, the first 8 in the sequence of the primer were used for testing. For NSP the 8 letters were used, which appeared most frequently in the instruction material (either the primer or the teachers' sight word lists). To assure that all children were familiar with the letters used in the reading tasks, this was checked by presenting a list of the 8 letters prior to reading assessment. If a child failed to name a letter, the letter was repeated by the researcher combined with a word example of its use. This procedure was repeated until every child was able to name all letters. For the word reading test, the percentage of correct answers was scored. Reading speed was not measured at this point.

June grade one and four reading: Children were given lists of 30 high and 30 low frequency words, and 30 nonwords (see Klicpera & Gasteiger-Klicpera, 1994). German words of different frequency were identified on the basis of "The German Basic Vocabulary" (Plickat, 1983). According to the CELEX word frequency database (see Baayen, Piepenbrock & Gulikers, 1995), high frequency and low frequency words have mean log frequencies per million of 1.86 and 1.38, respectively. The length of the words varied from one to three syllables. The words were presented in blocks of 15 in the same format as in grade one in November. For every list the number of correctly read words was counted and the reading speed was measured. The number of words per minute was calculated.

Teacher interviews: To get more information about the used reading instruction, teachers (n=49) were interviewed in the beginning of grade one. Various aspects of the teachers' approach to reading instruction were covered, e.g. investment of time, reading exercises, additional materials, primers used. The interview lasted about 30 minutes.

Testing took place in a quiet room in the school building during school hours. Children were taken out of class and tested face to face.

Statistical analysis

As argued above, the prediction of reading might be influenced by a different course of early reading development in SP and NSP classes. To demonstrate differences in early reading achievement ANOVAs with "type of word" (textbook/high frequent words, non-textbook/low frequent words, nonwords) were conducted to compare reading scores in SP and NSP for children "at risk" and "not at risk" according to the BISC separately for every reading measure.

For the analysis of the predictive strength of the BISC we used both correlational and classificatory analysis. A high correlation between the BISC and reading does not necessarily imply that low BISC scores predict a failure in the acquisition of reading. Only very high BISC scores might go hand in hand with high reading scores, yet children who score low on the predictor might still be able to catch up after the beginning of reading instruction. Therefore it is important not only to consider the broad range of reading achievement, but also to carry out classificatory analysis to see how successful the BISC is in identifying individual *reading problems*.

The classificatory analysis is based on the classification of children either "not at risk" or "at risk" according to the procedure given by the BISC manual, which was the predictor. A child was classified as "poor reader" when its reading score (all lists combined) fell below the 15% of all children within the instruction group (i.e. NSP or SP). 2 x 2 contingency tables were conducted and an index for the relative increase of the hit rate compared to the hit rate expected by chance (RATZ) was calculated. The index is essentially given by (hit rate – hit rate $_{\text{expected by chance}}$) / (maximal hit rate – hit rate $_{\text{expected by chance}}$). It is frequently used as an index for the quality of prediction (see Jansen et al., 1999). A value of 66% and more is considered an indicator of good prognostic validity, a value within the range of 34% - 66% is considered an admissible (although not specific) classification, and a value below 34% is unacceptable.

To analyze the relative importance of the predictors we conducted a series of simple structural equation models (SEM) with PA, RAN and VIS factor scores and the standardized SPM scores as reading predictors. The analysis was conducted in the following manner: In a first step we ensured that the correlations among the 3 BISC predictors and the intelligence were essentially the same for SP and NSP. In a confirmatory factor analysis we specified two different two-group models (i.e. the parameters were estimated simultaneously for NSP and SP in each of the two models). In model A, the correlations among the predictors were assumed as equal for SP and NSP, and in model B they were allowed to differ. These two models did not differ significantly ($\Delta\chi^2_{df=3} = 2.3$). Hence, the correlations among the predictors were fixed at the values of model A for subsequent analyses.

In a second step we added reading measures (each measure separately) to the equation as dependent variables (figure 1). In each case we compared a full-restricted model (all predictor effects equal) to a semi-restricted model (only restrictions on predictor correlations)

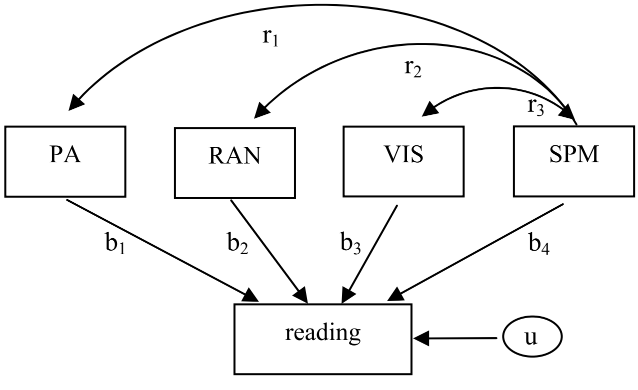

Figure 1:
SEM for the relative contribution of PA, RAN, VIS and intelligence (SPM) to reading. Semi-restricted model: All predictor correlations (r_i) specified as being the same for SP and NSP. Full-restricted model: Predictor correlations and coefficients b_1 to b_4 specified as being the same for SP and NSP. Unique variance of reading is denoted as "u".

resulting in a total of 15 comparisons (table 6). If for a given reading measure the full-restricted model differed significantly from the semi-restricted model according to the overall χ^2-test we used critical ratios for differences (CR; e.g. Byrne, 2001) to interpret parameter differences for the single predictors.

The analysis was conducted using SPSS® for Windows ver. 15 and AMOS® 7 for structural equation modelling. The level of significance was set to $\alpha=.05$ for all analysis.

Results

Descriptive statistics and course of reading development

Reading instruction in SP and NSP classes: No differences were found regarding age of teacher, years of experience in teaching, and number of students per class (table 2). Marginal differences were shown for implementing additional materials, homework and reading related exercises. As expected, more SP teachers used primers in everyday reading instruction. Interestingly they also used additional materials (e.g. from online resources) to a greater extent than NSP teachers. However, no single NSP teacher was found who did *not* use self-prepared materials (mostly worksheets with sight word lists). This was the case for 50% of the SP teachers.

Results for additional reading related exercises were somewhat unclear. More distinct differences were found at written (but not verbal) phoneme substitution exercises and verbal phoneme isolation exercises, both being more practiced in SP classes. On the other hand, "alternative" exercises like for example phoneme pronunciation trainings in front of a mirror appeared more often in NSP classes. In terms of homework NSP teachers handed out shorter (but not fewer) exercises than their SP colleagues.

Predictors: Children attending NSP classes showed slightly higher intelligence scores ($p<.05$). There were no significant differences between the later NSP and SP groups in terms of phonological awareness, naming speed, and visual word recognition (table 3). In the SP group 15.2% of the children were identified "at risk" according to the BISC manual, and 14.3% in the NSP group ($p>.05$). These percentages are somewhat higher than reported by Marx and Weber (2006; 9.1%) and Jansen et al. (1999: 12.4%).

Development of reading accuracy: After three months of reading instruction, remarkable differences were found between NSP and SP. Generally, children in SP classes were more accurate in reading unfamiliar words and nonwords than their peers in NSP classes. This held both for children "not at risk" as well as children classified as being "at risk," although differences were more pronounced for the latter group (table 3). Children "at risk" in NSP classes, on average, correctly read only about 35% of non-textbook words and 20% of nonwords. In the SP group, children "at risk" approached their "no risk" peers. The former group also read about 60% correctly, thus reading at approximately the level of the "no risk" children in the NSP group. Results of a RISK (risk, no risk) × INSTRUCTION (NSP, SP) × TYPE (textbook/high frequent, non-textbook/low frequent, nonword) repeated measurement ANOVA (table 4) showed a significant main effect for RISK, INSTRUCTION, and TYPE. Generally, children "at risk" and children in NSP classes read fewer words correctly. Furthermore all children made more errors in non-textbook words and nonwords. Interestingly,

Table 2:
Teacher characteristics and reading instruction in NSP and SP classes

	NSP			SP		
Teachers and students						
age of teacher in years (mean, std.)	42.2 (12.2)			42.7 (8.5)		
years of teacher in school (mean, std.)	20.0 (12.0)			20.1 (9.2)		
students in class (mean, std.)	19.5 (2.1)			19.6 (3.9)		
Primer (% of teachers)						
use of primer (every day, 3-4 times/week, less than 3 times/week)	42.9	28.6	28.6	65.0	25.0	10.0
how many lessons out of primer (all, about 2/3, about 1/4 or less)	16.7	66.7	16.7	100	-	-
Homework (% of teachers)						
frequency of homework (every day, 4 times/week, less than 4 times/week)	87.5	12.5	-	90.0	05.0	05.0
estimated duration of homework (10-20 min., 21-30 min., > than 30 min.)	62.5	37.5	-	25.0	65.0	10.0
Source of homework						
primer (% of agreement)	75.0			80.0		
worksheets (webbased resources) (% of agreement)	62.5			85.0		
teacher (% of agreement)	37.5			50.0		
Additional materials						
worksheets from primer (% of agreement)	50.0			65.0		
worksheets (e.g. from the internet, % of agreement)	37.5			75.0		
self-prepared worksheets (% of agreement)	100			50.0		
Additional reading exercises						
phoneme substitution, verbal (% of agreement)	62.5			55.0		
phoneme substitution, written (% of agreement)	25.0			80.0		
phoneme isolation, verbal (% of agreement)	50.0			75.0		
phoneme training, mirror (% of agreement)	87.7			15.8		
phoneme blending (% of agreement)	62.5			47.5		

the interaction terms TYPE × INSTRUCTION and TYPE × RISK were significant, indicating that both, children in NSP classes and children “at risk”, had more problems with unfamiliar words and nonwords. However, although the RISK × INSTRUCTION interaction was not significant, there was a significant RISK × INSTRUCTION × TYPE triple interaction. Post hoc tests showed that differences between NSP and SP were limited to unfamiliar words and nonwords. For familiar textbook words no differences in reading accuracy between NSP and SP were found.

The subsequent development of reading accuracy (June grades one and four) can be summed up as follows (tables 3 and 4). The main effect for TYPE remains significant in grades one and four, indicating that frequent words were easier to decode than unfrequent words and nonwords. The main effect of RISK remained significant until the end of grade one, indicating that children "at risk" read, on average, fewer words correctly than their "no risk" peers. The INSTRUCTION × RISK interaction at the end of grade one was significant. However, the effect size for this effect was low ($\eta^2 = .02$). Importantly, no significant main effect for INSTRUCTION was found at the end of grades one or four; there was also no 2-way or triple interactions involving reading instruction (with the exception of the one reported above).

Table 3:
Predictors (T-values; for intelligence SPM raw scores) and reading measures for SP and NSP. Means and Std

predictors		NSP		SP	
	Intelligence	27.2	(10.2)	**29.4**	(9.7)
	PA	48.5	(10.5)	50.3	(10.1)
	RAN	49.6	(9.9)	50.2	(10.1)
	VIS	48.5	(8.8)	50.1	(10.4)
reading measures		NSP		SP	
		risk	no risk	risk	no risk
November, grade 1					
% correct	textbook	85.0 (14.1)	94.1 (8.0)	85.7 (17.3)	92.2 (11.5)
	non textbook	34.7 (28.3)	52.6 (32.4)	59.7 (32.5)	74.9 (26.6)
	nonwords	20.0 (22.9)	52.0 (33.5)	61.0 (33.6)	77.1 (24.1)
June, grade 1					
% correct	high frequency	77.1 (17.0)	89.0 (7.8)	82.4 (12.8)	86.4 (12.6)
	low frequency	77.6 (17.0)	88.2 (7.8)	80.4 (15.0)	84.5 (14.0)
	nonwords	81.0 (21.6)	84.6 (11.5)	79.4 (15.9)	81.7 (15.0)
words per minute	high frequency	20.7 (8.2)	19.3 (7.0)	22.6 (12.0)	23.7 (11.4)
	low frequency	19.1 (7.8)	17.7 (6.8)	20.2 (10.1)	21.0 (9.2)
	nonwords	18.3 (6.8)	17.1 (6.0)	19.4 (8.3)	19.6 (7.9)
June, grade 4					
% correct	high frequency	96.7 (4.2)	97.9 (2.2)	97.4 (3.5)	97.5 (5.1)
	low frequency	94.0 (6.3)	94.1 (4.8)	94.9 (4.9)	95.4 (5.7)
	nonwords	85.1 (8.8)	87.0 (8.7)	88.9 (8.6)	88.7 (10.6)
words per minute	high frequency	71.9 (15.5)	83.4 (20.5)	71.8 (20.1)	87.2 (24.2)
	low frequency	55.9 (16.3)	65.9 (21.4)	57.2 (17.4)	70.3 (23.1)
	nonwords	38.3 (8.9)	43.1 (12.7)	38.8 (9.7)	43.7 (13.7)

Table 4:
ANOVA summary for the development of reading skills. Comparison of children identified “at risk” and “not at risk” according to the BISC in SP and NSP classes. Bold values denote significant effects

	% correct						Words per minute			
	Nov., grade 1		June, grade 1		June, grade 4		June, grade 1		June, grade 4	
	F	η^2	F	η^2	F	η^2	F	η^2	F	η^2
instruction	**47.89**	.12	.23	.00	3.14	.01	1.67	.01	.53	.01
risk	**35.73**	.09	**19.69**	.05	.64	.00	**7.99**	.02	**17.39**	.05
type	**380.62**	.52	**20.86**	.06	**204.60**	.40	**41.49**	.11	**731.55**	.70
instruction × risk	1.67	.01	**5.55**	.02	.42	.01	.15	.01	.24	.01
instruction × type	**68.96**	.17	1.41	.01	3.11	.01	2.39	.01	.71	.01
type × risk	**15.17**	.04	.61	.01	.13	.01	**7.43**	.02	**10.35**	.03
type × risk × instruction	**3.38**	.01	.34	.01	.73	.01	.29	.01	.54	.01

Development of reading speed: For reading speed, our findings were similar. ANOVAs resulted in significant effects for RISK and TYPE as well as RISK × TYPE in grades one and four. Children “at risk” read slower than their “no risk” peers. This difference was larger for unfrequent words and nonwords than for frequent words (which were generally more difficult to read). However, as with reading accuracy, no statistically reliable effects involving INSTRUCTION were found.

Prediction of reading problems

Before looking at the relative contribution of the BISC factors for the prediction of reading, we will report results concerning the prognostic validity of the instrument. Table 5 shows the results of the classificatory analysis. For *reading accuracy* RATZ-indices were higher for NSP than for SP on all measures, indicating a satisfying prognostic validity for grade one test dates (both, November and June) for NSP. For SP a marginal prediction was found for November, grade one. The difference was not so much in risk affirmation (i.e. the percentage of children being classified “at risk” and having developed reading problems later; it was low both for NSP and SP), but for sensitivity. For instance, less than half of the children having reading problems at the end of grade one in SP classes could be identified at kindergarten age. The respective percentage for NSP was 80%. For *reading speed*, none of the RATZ indices reached a satisfying level, and both risk affirmation and sensitivity were low.

Table 5:
Risk affirmation and sensitivity of the BISC, RATZ-indices, and correlations between BISC risk-points and reading measures. Bold values denote significant differences of the correlation coefficients in SP and NSP. All correlations are significant except SP reading accuracy in grades one and four (r = -.11; r = -.01) and NSP reading speed in June grade one (r = -.18)

	November, grade 1		June, grade 1		June, grade 4	
	NSP	SP	NSP	SP	NSP	SP
Reading accuracy						
% risk affirmation	23.3	28.0	26.7	23.1	19.2	14.5
% sensitivity	77.8	59.0	80.0	47.5	50.0	30.3
RATZ	67.0	40.0	70.3	23.3	27.6	0.1
r (risk points)	**-.43**	**-.24**	**-.42**	**-.11**	**-.23**	**-.01**
Reading speed						
% risk affirmation			16.7	18.3	23.1	25.0
% sensitivity			38.5	38.5	40.0	45.9
RATZ			8.7	8.5	13.1	22.7
r (risk points)			-.18	-.16	-.33	-.18

Contribution of PA, RAN and VIS to the prediction of reading

Reading accuracy: SEM analysis (table 6) showed remarkable differences between NSP and SP in the prediction of reading three months after the beginning of reading instruction. Multiple R^2s were lower for SP than for NSP for most measures of reading *accuracy*. Furthermore, the full restricted models (i.e. equal parameters for NSP and SP) each had a fairly good fit. However, the predictive strength of the three BISC predictors differed significantly in terms of global model fit (i.e. significant differences in favour of the semi-restricted model which allowed the regression weights to be different for NSP and SP) for textbook words and nonwords. The observed pattern can be described as follows: (1) Children in NSP classes generally relied more on their early visual skills than their peers in SP classes. For NSP, approximately 13% of the variance in textbook word reading and 6% of nonword reading was explained by VIS. For SP, the respective percentages were close to zero for both textbook words and nonwords. (2) Children in NSP classes relied more heavily on their early phonological skills when nonwords were used as reading measures (although PA is also a significant predictor for SP). The same trend, although not statistically significant, was found for non textbook words. For highly familiar textbook words, PA predicted both NSP and SP about equally. (3) For children in SP (but not NSP) classes, RAN was found to be a significant predictor for both types of (real) words, while the reverse pattern was found for nonwords; RAN was a predictor for NSP but not for SP, although the critical ratio of difference (CR_{diff}) was not significant. However, for non textbook words, parameter differences between NSP and SP in terms of global model fit were not significant, although the pattern was roughly the same as for textbook words.

Table 6:

SEM analysis (standardized solution) for the contribution of phonological awareness (PA), naming speed (RAN), and visual word matching (VIS) to the prediction of reading: Parameters, critical ratios (values in brackets), critical ratios parameter differences (CRdiff), FIT-indices (χ^2, AGFI, RMSEA), and model comparison ($\Delta\ \chi^2$). df (full-restricted model) = 13; df (semi-restricted model) = 9. Significant values in bold letters

		reading accuracy								
		Textbook / high frequency			non textbook/low frequency			nonwords		
		NSP	SP	CRdiff	NSP	SP	CRdiff	NSP	SP	CRdiff
November, grade 1	SPM	.06 (0.5)	.16 **(2.3)**	-1.30	.15 (1.3)	.28 **(4.1)**	-1.05	.17 (1.6)	.29 **(4.3)**	-.90
	PA	.23 **(2.3)**	.15 **(2.4)**	-.05	.27 **(2.5)**	.11 (1.7)	1.36	.40 **(4.3)**	.13 **(2.1)**	**2.67**
	RAN	.01 (0.0)	.20 **(2.9)**	-1.87	-.05(-0.4)	.14 **(2.2)**	-1.33	.20 **(2.0)**	.12 (1.9)	.91
	VIS	.36 **(3.5)**	.01 (0.2)	**2.62**	.12 (1.1)	.02 (0.4)	.83	.24 **(2.6)**	.05 (0.7)	**2.03**
	multiple R²	.20	.12		.12	.15		.34	.16	
	full-restricted model:	χ^2= 18.9; AGFI=.95 ; RMSEA=.04			χ^2= 11.9; AGFI=.97; RMSEA=.00			χ^2= 17.0; AGFI=.95; RMSEA=.03		
	semi-restricted model:	χ^2=5.3; AGFI=.98 ; RMSEA=.00			χ^2= 5.3; AGFI= .98; RMSEA=.00			χ^2= 5.3; AGFI=.98; RMSEA=.00		
		$\Delta\ \chi^2$ = **13.65**			$\Delta\ \chi^2$ = 6.64			$\Delta\ \chi^2$ = **11.75**		
June, grade 1	SPM	-.02(-0.2)	.12 (1.7)	-1.20	-.13(-1.1)	.11 (1.6)	-1.84	.06 (0.5)	.08 (1.1)	-.25
	PA	.26 **(2.5)**	.13 **(2.1)**	.06	.19 (1.7)	.11 (1.6)	.30	.09 (0.8)	.14 **(2.1)**	-.45
	RAN	.13 (1.1)	.18 **(2.6)**	-.60	.17 (1.4)	.11 (1.6)	.25	-.01(-0.1)	.08 (1.1)	-.58
	VIS	.28 **(2.6)**	.04 (0.5)	1.92	.26 **(2.4)**	.05 (0.7)	1.60	.17 (1.5)	.01 (0.2)	1.27
	multiple R²	.16	.09		.12	.05		.05	.04	
	full-restricted model:	χ^2= 11.3; AGFI=.97; RMSEA=.00			χ^2= 10.5; AGFI=.97; RMSEA=.00			χ^2= 7.5; AGFI=.98; RMSEA=.00		
	semi-restricted model:	χ^2= 5.3; AGFI=.98; RMSEA=.00			χ^2= 5.3; AGFI=.98; RMSEA=.00			χ^2= 5,3; AGFI=.98; RMSEA=.00		
		$\Delta\ \chi^2$ = 6.07			$\Delta\ \chi^2$ = 5.24			$\Delta\ \chi^2$ = 2.25		
June grade 4	SPM	.03 (0.2)	.16 **(2.2)**	-1.48	.11 (0.9)	.25 **(3.6)**	-1.18	.28 **(2.3)**	.19 **(2.6)**	.30
	PA	.10 (0.9)	-.01(-0.1)	.71	.16 (1.4)	-.05(-0.8)	1.66	.11 (1.0)	-.09(-1.4)	.16
	RAN	.18 (1.5)	.10 (1.4)	.28	.13 (1.0)	.11 (1.7)	.15	-.09(-0.7)	.02 (0.3)	-.72
	VIS	.16 (1.5)	-.06(-0.9)	1.68	-.01(-0.1)	-.07(-1.1)	.41	-.08(-0.7)	.01 (1.0)	-.64
	multiple R²	.08	.05		.07	.10		.09	.04	
	full-restricted model:	χ^2= 9.9; AGFI=.97; RMSEA=.00			χ^2= 9.2; AGFI=.97; RMSEA=.00			χ^2= 8.5; AGFI=.97; RMSEA=.00		
	semi-restricted model:	χ^2= 5.2; AGFI=.98; RMSEA=.00			χ^2= 5.3; AGFI=.98; RMSEA=.00			χ^2= 5.2; AGFI=.98; RMSEA=.00		
		$\Delta\ \chi^2$ = 4.68			$\Delta\ \chi^2$ = 3.93			$\Delta\ \chi^2$ = 3.36		

continued

		reading speed								
		high frequency			low frequency			nonwords		
		NSP	SP	CRdiff	NSP	SP	CRdiff	NSP	SP	CRdiff
June, grade 1	SPM	.09 (0.8)	.13 (1.9)	-.63	.13 (1.0)	.15 **(2.2)**	-.56	.11 (0.9)	.11 (1.6)	-.29
	PA	.10 (0.8)	.10 (1.6)	-.30	.02 (0.2)	.08 (1.2)	-.56	-.08(-0.7)	.03 (0.4)	-.82
	RAN	.16 (1.3)	.27 **(4.0)**	-1.18	.13 (1.1)	.25 **(3.6)**	-1.10	.11 (0.9)	.30 **(4.4)**	-1.64
	VIS	.11 (0.9)	-.08(-1.3)	1.49	.14 (1.2)	-.06(-0.9)	1.50	.09 (0.8)	-.08(-1.3)	1.34
	multiple R^2	.07	.13		.07	.12		.05	.13	
	full-restricted model:	χ^2= 9.7; AGFI=.97; RMSEA=.01			χ^2= 9.5; AGFI=.97; RMSEA=.01			χ^2= 10.8; AGFI=.97; RMSEA=.01		
	semi-restricted model:	χ^2= 5.0; AGFI=.98; RMSEA=.00			χ^2= 5.0; AGFI=.98; RMSEA=.00			χ^2= 5.0; AGFI=.98; RMSEA=.00		
		$\Delta\chi^2 = 4.73$			$\Delta\chi^2 = 4.53$			$\Delta\chi^2 = 5.79$		
June, grade 4	SPM	-.04(-0.4)	.25 **(3.6)**	**-2.56**	-.03(-0.2)	.20 **(2.9)**	-1.86	-.03(-0.3)	.10 (1.5)	-1.08
	PA	.07 (0.6)	.07 (1.1)	-.19	.15 (1.4)	.08 (1.3)	.45	.03 (0.3)	-.03(-0.5)	.51
	RAN	.22 (1.9)	.21 **(3.2)**	-.19	.22 (1.8)	.27 **(4.0)**	-.42	.21 (1.8)	.22 **(3.2)**	-.09
	VIS	.30 **(2.8)**	-.02(-0.2)	**2.49**	.16 (1.4)	-.01(-0.1)	1.27	.20 (1.7)	.04 (0.6)	1.25
	multiple R^2	.13	.15		.08	.16		.07	.08	
	full-restricted model:	χ^2= 16.8; AGFI=.95; RMSEA=.03			χ^2= 10.8; AGFI=.97; RMSEA=.02			χ^2= 7.8; AGFI=.98; RMSEA=.01		
	semi-restricted model:	χ^2= 4.9; AGFI=.98; RMSEA=.00			χ^2= 5.0; AGFI=.98; RMSEA=.00			χ^2= 5.0; AGFI=.98; RMSEA=.01		
		$\Delta\chi^2 = \mathbf{11.88}$			$\Delta\chi^2 = 5.86$			$\Delta\chi^2 = 2.84$		

For the later stages of reading achievement (i.e. end of grades one and four), no significant differences in the predictors were found according to the comparison of the global model fit for restricted and semi-restricted models, and according to the CR_{diff}. However, VIS appeared by trend to be a better predictor for all reading measures in NSP than in SP classes.

Reading speed: For reading speed, with one exception (see below), no significant differences in the contribution of single BISC predictors could be found in terms of global model fit and CR_{diff}. Nevertheless, two sets of results are still worth mentioning. First, for all measures of reading speed, RAN was a better predictor than PA. It also appeared to be the only BISC factor that contributed to the prediction of reading. In the SP group, RAN was significant for all measures. Second, for VIS, the same tendency was found as with reading accuracy: Children in NSP classes tended to rely more on early visual pattern matching competencies than their peers in SP classes. For high frequency words in grade four, this difference was significant according to model comparison and CR_{diff}.

Summary and discussion

The aim of this study was to explore the influence of instruction on the prediction of reading by comparing extreme groups based on the material teachers used. We hypothesized that reading skills, and hence the prediction of reading, are strongly influenced by the systematics of letter-to-sound instruction at early stages of reading achievement. In support of this hypothesis we found the following main results:

1) When *reading accuracy* was taken as measure, the multiple R^2s were considerably higher for NSP than for SP, particularly when the demands on phonological decoding were high (i.e. nonwords). When the demands on phonological decoding were low (or when lexical reading was possible) children in NSP classes relied on their visual skills (i.e. on a visual word matching strategy).
2) The prediction of individual reading problems (classificatory analysis), with reading accuracy as criterion, was consistently better for NSP than for SP.
3) When *reading speed* was taken as measure, there were hardly any reliable differences in the predictive strength of the three components PA, RAN, and VIS between SP and NSP. Yet, NSP tended more to rely on visual word matching strategies.

The interpretation of this pattern is relatively clear if the development of reading *accuracy* in the group of children "at risk" in SP and NSP is considered. Our results showed clearly that in SP classes, children with initial little PA had a better chance to catch up and even develop appropriate phonological skills via reading and reading instruction than their peers in NSP. In fact, children "at risk" according to the Bielefelder Screening read on average 61% nonwords correctly in November of grade one in the SP group, while the percentage in NSP was only 20%. Further support to the idea, that phonological skills are primarily "instruction-induced" emerged from an additional analysis. In a subsample, children "at risk" in SP performed even better than children "not at risk" in NSP on a phoneme substitution task in November of grade one. The latter accomplished 47% of the items, while the average score for the SP risk group was 53% (with the NSP risk group accomplishing only 26% of the items). On the other hand, when the demands on phonological recoding were lower, no differences in the reading scores between NSP and SP were found. Children in NSP relied more on their visual skills to “identify” target words. From this pattern, we concluded a strong influence of the reading instruction on phonological and probably none on visual components (although the predictive strength of VIS was considerably lower for SP than NSP). However, although we cannot provide any direct measures of VIS in grade one or later, we believe that rather than developing “instruction induced” visual skills, children in SP do simply not rely on a visual word matching strategy in favour of more efficient decoding strategies. Concerning RAN, it is interesting that this measure was predictive for words in SP but not in NSP in grade one in November. In our point of view, this finding is best interpreted in the way that (1) RAN is not touched by reading instruction (and therefore might not reflect orthographic processing) and (2) RAN influences reading primarily if basic phonological competencies are already developed. The first point can also be supported by the results for reading speed (see below); the second is consistent with the self-teaching hypothesis (Jorm & Share, 1983) where basic decoding skills are assumed as “starting point” for skilled automatic reading. Furthermore, RAN becomes increasingly important for word-reading over time, particularly in the NSP group, i.e. when – after a delay of several months – appropriate decoding skills are finally developed in this group.

From all our findings it is obvious that differences between SP and NSP in prediction of reading accuracy do not persist: Until the end of grade one, children in NSP caught up in their decoding competencies. We argued that this might be a consequence of the simplicity with which letter-to-sound mapping skills can be acquired in relatively regular orthographies even without direct and systematic instruction. Combined with this, a self teaching mechanism sensu Jorm and Share (1983) could be induced just by reading experience. However, this process may take somewhat longer for poor readers and readers “at risk”. In our study no instruction effects were found in the long-run.

For reading *speed*, no differences between NSP and SP were found concerning the predictive strength both of PA and RAN. With reading accuracy, children in NSP tended to rely more on their visual word matching skills.

RAN was consistently a predictor, but not PA. This is consistent with the results reported by Landerl and Wimmer (2008) and earlier findings in transparent orthographies (but different from most findings in phonologically opaque orthographies such as English). In these studies, PA was only a predictor for early stages of reading instruction, while RAN remained important, when reading speed was used as criterion. Landerl and Wimmer (2008) discussed whether this pattern appeared due to the fact that a new PA task was used in their study where participants did not have to segment phonemes independently, but had to imitate a segmentation modelled by the experimenter. However, by using factor scores of four quite traditional PA tasks we found similar results.

Altogether, we can confirm the position that in regular orthographies PA is rather an indicator of the children’s acquisition of the alphabetic principle than a prerequisite (see also Landerl & Wimmer, 2000). The interesting point however is how far PA is still important considering the easiness with which it can be acquired once reading instruction has begun – this might be of predictive strength. A detailed answer to this question must be subject of further research, but the analysis of a subsample of our children, for whom PA measures were conducted at the beginning and end of grade one, demonstrated that November grade one PA is uniquely predictive both for reading accuracy and reading speed at the end of grade one, even if early (November grade one) reading accuracy is controlled for. This effect appeared for both NSP and SP, but was remarkably stronger for reading accuracy in NSP (NSP: ΔR^2=.19; SP: ΔR^2=.04). However, June grade one PA does not contribute any more to reading accuracy in grade two when June grade one reading accuracy is partialled out. From this point of view we can confirm the assumptions that the simplicity of PA-acquisition might be of predictive value for the early stages of learning to read.

To sum up, we must conclude that there was only little success in predicting later reading achievement when using measures of the BISC. Only under specific circumstances, i.e. if there was no systematic letter-to-sound reading instruction, and only at the beginning of reading instruction, the analysis indicated a satisfying prognostic validity with RATZ-indices similar to the ones reported by Jansen et al. (1999). However, in SP the percentage of explained variance in reading accuracy was no more than 12-16% at the beginning of reading instruction, and for both groups it was 8-16% after 4 years of school. But even when reading speed was used as measurement (which might be more successful in measuring reading achievement as German speaking children read close to ceiling when looking at reading accuracy; Klicpera & Schabmann, 1993), similar low percentages (and even lower) were found. Our results on the SP group seem to confirm recent findings concerning the prognostic validity of the BISC. For instance, Marx and Weber (2006) reported RATZ-indices of 33% in grade one and 16% in grade four, for reading *speed*. Our findings complement these results by helping to clarify the role of reading instruction.

We argued from a theoretical perspective that the combination of a systematic phonic instruction *and* a shallow orthography would make prediction difficult. Thus, although our study only looked at German speaking children, we believe that our results concerning the influence of reading instruction on the prediction of reading skills might be generalizable to other transparent orthographies such as Dutch, Greek, Spanish and Italian (see e.g. Seymour, Aro & Erskine, 2003 for an overview of orthographic regularity). It is, however, difficult to value our results from a more "practical" perspective. On the one hand, children need good reading instruction, especially initially poor readers who must catch up. From this point of view, our results seem satisfying. On the other hand the early detection of children "at risk" is important in terms of early intervention, and our findings are clearly discouraging in this regard. From our theoretical perspective we must conclude, that sufficiently reliable predictors are thus far not available. We hence think that it might be more promising to predict reading success or reading failure as soon as possible *after the beginning of reading instruction*. We believe that currently a reliable identification of children at risk shortly after the beginning of reading instruction will be more helpful to most children than an unreliable one in kindergarten. Also, it might be easier to incorporate an early screening in school rather then kindergarten as reading still is a "task that is learned in school". For an effort to develop a method of early testing (three months after the beginning of reading instruction) see Klicpera, Humer, Gasteiger-Klicpera and Schabmann (2008).

Finally we must address some noteworthy limitations in our study. First, due to the longitudinal design of our study that was carried out as a field study taking place in schools, only extreme groups could be compared. NSP and SP are unequally distributed in our sample, but this may reflect the proportion in the population (although no exact data about this issue are available to our best knowledge). It should clarified at this point that the objective of our study was not to discuss the advantages or disadvantages of SP and NSP respectively; instead, we aimed to provide explanations of why the prediction of reading (particularly from early PA measures) is difficult when a shallow orthography and a systematic letter-to-sound reading instruction are involved (e.g. Marx and Weber, 2006). From this point of view, the fact that NSP is by far the less frequently used method of reading instruction is most likely of minor importance for our findings.

Even though we believe that our classification was well justified, there might be instructional markers which we could not assess appropriately. We do not know *exactly* to which extent additionally letter-to-sound exercises were given in our classes and how they looked like in detail. Our questionnaire could only roughly cover all activities during reading instruction. However, as mentioned above, the criteria used for the identification of SP and NSP seem feasible according to earlier work (e.g. Schabmann, 2007). Furthermore, the results presented for the early stages of reading acquisition (November grade one) justify the rationale of our method. The observed pattern of reading development is theoretically well-grounded, and shows remarkable and well interpretable differences between SP and NSP.

Perhaps more important from a "practical" perspective is that we only assessed predictor tasks that were used in the BISC. Therefore, our results might not be generalizable to other tasks, particularly different PA tasks as the ones used by Landerl and Wimmer (2008) which perhaps are closer to the underlying construct of PA and easier for preschoolers. On the other hand, most of the tasks are theoretically established and repeatedly used in literature. Hence, we strongly believe that our results show a more general problem (i.e. the prediction of possible reading success and failure before the beginning of reading instruction) rather than problems which are linked to the particular instrument which we used in this study.

References

Arbuckle, J. L. (2006). *Amos 7.0* [Computer Software]. Chicago: Smallwater.

Baayen, R. H., Piepenbrock, R., & Gulikers, L. (1995). *The CELEX lexical database* (CD-ROM). Philadelphia, PA: Linguistic Data Consortium, University of Pennsylvania.

Bishop, A. G. (2003). Prediction of first-grade reading achievement: A comparison of fall and winter kindergarden screenings. *Learning Disability Quarterly, 26,* 189-200.

Blachmann, B. A. (1984). Relationship of rapid naming ability and language analysis skills to kindergarten and first-grade reading achievement. *Journal of Educational Psychology*, *76,* 610-622.

Bowers, P. G., & Swanson, L. B. (1991). Naming speed deficits in reading disability: Multiple measures of a singular process. *Journal of Experimental Child Psychology, 51,* 195-219.

Bowers, P. G., & Wolf, M. (1993). Theoretical links among naming speed, precise timing mechanisms and orthographic skill in dyslexia. *Reading and Writin*g, 5, 69-85.

Bowers, P. G. (1995). Tracing symbol naming speed's unique contributions to reading disabilities over time. *Reading and Writing: An Interdisciplinary Journal, 7,* 189-216.

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

Bryant, P., Bradley, L., MacLean, M., & Crossland, J. (1989). Nursery rhymes, phonological skills and reading. *Journal of Child Language, 16,* 407-428.

Byrne, B. M. (2001). *Structural equation modelling with AMOS: Basic concepts, applications, and programming*. New Jersey: Lawrence Erlbaum Association.

Caravolas, M., Volín, J., & Hulme, C. (2005). Phoneme awareness is a key component of alphabetic literacy skills in consistent and less consistent orthographies: Evidence from Czech and English children. *Journal of Experimental Child Psychology, 92,* 107-139.

Compton, D. L., DeFries, J. C., & Olson, R. K. (2001). Are RAN- and phonological awareness-deficits additive in children with reading disabilities? *Dyslexia, 7,* 125-129.

Cornwall, A. (1992). The relationship of phonological awareness, rapid naming, und verbal memory to severe reading and spelling disability. *Journal of Learning Disabilities, 25,* 532-538.

Cossu, G., Shankweiler, D., Liberman, I. Y., Katz, L., & Tola, G. (1988). Awareness of phonological segments and reading-ability in Italian children. *Applied Psycholinguistics, 9,* 1-16.

Ehri, L. C. (1993). Commentary on Marx, Jansen Mannhaupt and Skowronek: The Bielefeld Screening Battery to predict reading and spelling difficulties. In H. Grimm & H. Skowronek (Eds.), *Language acquisition problems and reading disorders: Aspects of diagnosis and intervention* (pp. 234-255). Berlin: Walter de Gruyter.

Ferdinand, W. (1970). *Über die Erfolge des ganzheitlichen und synthetischen Lese-(Schreib)unterrichts in der Grundschule* (Achievements of whole word and synthetic writing instruction in elementary school). Verlag Neue Deutsche Schule, Essen.

Georgiou, G. K., Parrila, R., Kirby, J. R., & Stephanson, K. (2008). Rapid naming components and their relationship with phonological awareness, orthographic knowledge, speed of processing, and different reading outcomes. *Scientific Studies of Reading, 12,* 325-350.

Holzinger, F. (1964). *Leistungserhebung auf der vierten Schulstufe* (Skill examination in fourth grade). Leykam, Graz and Vienna.

Jansen, H., Mannhaupt, G., Marx, H., & Skowronek, H. (1999). *Bielefelder Screening zur Früherkennung von Lese- und Rechtschreibschwierigkeiten* (The Bielefeld Screening Battery to predict reading and spelling difficulties) (BISC). Handanweisung (Testmanual). Göttingen: Hogrefe.

Jorm, A. F., & Share, D. L. (1983). Phonological recoding and reading acquisition. *Applied Psycholinguistics, 4,* 103-147.

Kail, R., Hall, L. K., & Caskey, B. J. (1999). Processing speed, exposure to print, and naming speed. *Applied Psycholinguistics, 20,* 303-314.

Kirby, J. R., Parrila, R., & Pfeiffer, S. L. (2003). Naming speed and phonological awareness as predictors of reading development. *Journal of Educational Psychology, 80,* 437-447.

Klicpera, C., & Schabmann, A. (1993). Do German-speaking children have a chance to overcome reading and spelling difficulties? *European Journal of Psychology of Education, 3,* 307-323.

Klicpera, C., & Gasteiger-Klicpera, B. (1994). Die langfristige Entwicklung der mündlichen Lesefertigkeit bei guten und schwachen Lesern (The longterm development of oral reading skills in good and poor achieving readers). *Zeitschrift für Entwicklungspsychologie und Pädagogische Psychologie, 26,* 278-290.

Klicpera, C., Humer, R., Gasteiger-Klicpera, B., & Schabmann, A. (2008). *WFT – Wiener Früherkennungstest.* Leselehrgangspezifisches Verfahren zur Früherkennung von Kindern mit Lese- und Rechtschreibschwierigkeiten im Schuleingangsbereich. (*Viennese Early Diagnosis Test.* A schoolbook based method for the early dedection of children with reading and spelling difficulties) Wien: Verlag E. Dorner.

Kratzmeier, H., & Horn, R. (1988). *RAVEN-Matrizen-Test. Standard Progressive Matrices (SPM)* (2. erweiterte und überarbeitete Aufl.). Weinheim: Beltz.

Landerl, K., & Wimmer, H. (2000). Deficits in phoneme segmentation are not the core problem of dyslexia: Evidence from German and English children. *Applied Psycholinguistics, 21,* 243-262.

Landerl, K., & Wimmer, H. (2008). Development of word reading fluency and spelling in a consistent orthography: An 8-year follow-up. *Journal of Educational Psychology, 100,* 150-161.

Manis, F. R., Doi, L. M., & Badha, B. (2000). Naming speed, phonological awareness, and orthographic knowledge in second graders. *Journal of Learning Disabilities, 33,* 325-333.

Marx, H. (1985). *Aufmerksamkeitsverhalten und Leseschwierigkeiten* (Attention and reading difficulties). Weinheim: VCH-Verlagsgesellschaft.

Marx, P., & Weber, J. (2006). Vorschulische Vorhersage von Lese- und Rechtschreibschwierigkeiten. (Preschool prediction of reading and spelling difficulties.) *Zeitschrift für Pädagogische Psychologie, 20*(4), 251-259.

McBride-Chang, C., & Manis, F. R. (1996). Structural invariance in the associations of naming speed. Phonological awareness, and verbal reasoning in good and poor readers: A test of the double deficit hypothesis. *Reading and Writing: An Interdisciplinary Journal, 8,* 323-339.

Moll, K., Fussenegger, B., Willbuger, 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.

Morais, J., Alegria, J., & Content, A. (1987). The relationship between segmental analysis and alphabetic literacy: An interactive view. *Cahiers de Psychologie Cognitive, 7,* 415-439.

Müller, H. (1964). *Methoden des Erstleseunterrichts und ihre Ergebnisse*. (Methods of elementary reading instruction and their results). Meisenheim.

Müller, R. (1965). Rechtschreibung und Fehleranalyse (Spelling and error analysis). *Schule und Psychologie, 12,* 161-173.

Neuhaus, G. F., & Swank, P. R. (2002). Understanding the Relations between RAN Letter Subtest Components and Word Reading in First-Grade Students. *Journal of Learning Disabilities, 35*(2), 158-174.

Parilla, R., Kirby, J. R., & McQuarrier, L. (2004). Articulation rate, naming speed, verbal short-term memory, and phonological awareness: Longitudinal predictors of early reading development? *Scientific Studies of Reading, 8,* 3-26.

Patel, T., Snowling, M., & de Jong, P. (2004). A cross-linguistic comparison of children learning to read in English and in Dutch. *Journal of Educational Psychology, 96,* 785-797.

Plickat, H. (1983). *Der deutsche Grundwortschatz* (German basic vocabulary). Weinheim: Beltz.

Schabmann, A., & Kabicher, R. (2007). Auswirkungen früher Leseschwierigkeiten – Leseverständnis, Leseverhalten und Einstellungen zum Lesen am Ende der Grundschulzeit. (The effects of reading difficulties: reading comprehension, reading bahaviour, and attitudes toward reading at the end of primary school) *Heilpädagogische Forschung, 23*(3), 155-166.

Schabmann, A. (2007). Erstleseunterricht und Lese-Rechtschreibleistungen – Ergebnisse einer Wiener Längsschnittuntersuchung. (Primary reading instruction and reading/spelling achievement – findings of a Viennese longitudinal study). In G. Schulte-Körne (Eds.), *Legasthenie und Dyskalkulie in Wissenschaft, Schule und Gesellschaft* (pp. 59-70). Verlag D. Winkle, Bochum.

Schatschneider, C., Francis, D. J., Carlson, C. D., Fletcher, J. M., & Foorman, B. R. (2004). Kindergarten prediction of reading skills: A longitudinal comparative analysis. *Journal of Educational Psychology, 96,* 265-282.

Schubenz, S. (1966). Soll am Beginn des Rechtschreib- (und Lese-) Unterrichts die Synthese oder die Analyse der Texteinheiten betont werden? (Should the synthesis or analysis of text units be emphasized at the beginning of spelling (and reading) instruction?) *Schule und Psychologie, 13,* 39-45.

Seymour, P. H. K., Aro, M., & Erskine, J. M. (2003). Foundation literacy acquisition in European orthographies. *British Journal of Psychology, 94,* 143-174.

Spring, C., & Davis, J. M. (1988). Relations of digit naming speed with three components of reading. *Applied Psycholinguistics, 9,* 315-334.

Torgesen, J. K., & Burges, S. R. (1998). Consistency of reading-related phonological processes throughout early childhood: Evidence from longitudinal-correlational and instructional studies. In J. Metsala & L. Ehri (Eds.), *Word recognition in beginning reading* (pp. 161-188). Hilsdale, NJ: Erlbaum.

Torgesen, J. K., Wagner, R. K., & Rashotte, C. A. (1998). *Test of word reading efficienc*y. Austin, TX: PRO-ED.

Verhagen, W., Aarnoutse, C., & van Leeuwe. J. (2006). Predictoren voor beginnende Geletterdheid [Prediction of early literacy]. *Pedagogische Studiën, 83,* 226-245.

Verhagen, W., Aarnoutse, C., & van Leeuwe, J. (2008). Phonological awareness and naming speed in prediction of Dutch children's word recognition. *Scientific Studies of Reading, 12*, 301-324.

Wimmer, H. (1993). Characteristics of developmental dyslexia in a regular writing system. *Applied Psycholinguistics, 14,* 1-22.

Wimmer, H., Landerl, K., Linortner, R., & Hummer, P. (1991). The relationship of phonemic awareness to reading acquisition. More consequence than precondition but still important. *Cognition, 40,* 219-249.

Wimmer, H., Mayringer, H., & Landerl, K. (2000). The double-deficit hypothesis and difficulties in learning to read a regular orthography. *Journal of Educational Psychology, 92,* 668-680.

Wolf, M., & Bowers, P. G. (1999). The double deficit hypothesis for the developmental dyslexia. *Journal of Educational Psychology, 91,* 415-438.

Wolf, M., Bowers, P. G., & Biddle, K. (2000). Naming-speed processes, timing and reading: A conceptional review. *Journal of Learning Disabilitie*s*, 33,* 387-407.