Reading comprehension tests and poor readers: How test processing demands result in different profiles

Timothy C. Papadopoulos, Panayiota Kendeou, Maria Shiakalli

Dans L'Année psychologique 2014/4 (Vol. 114), pages 725 à 752
Éditions NecPlus

ISSN 0003-5033
DOI 10.3917/anpsy.144.0725

Article disponible en ligne à l'adresse
https://www.cairn.info/revue-l-annee-psychologique1-2014-4-page-725.htm
Reading comprehension tests and poor readers: How test processing demands result in different profiles

Timothy C. Papadopoulos\textsuperscript{1}, Panayiota Kendeou\textsuperscript{2*} and Maria Shiakalli\textsuperscript{3}

\textsuperscript{1} Department of Psychology and Centre for Applied Neuroscience, University of Cyprus, Cyprus
\textsuperscript{2} Department of Educational Psychology, University of Minnesota, USA
\textsuperscript{3} Department of Psychology, University of Cyprus, Cyprus

ABSTRACT
This study investigated different subtypes of poor readers, following an original group of 213 children from kindergarten to Grade 2. Four groups were formed on the basis of their performance on three reading comprehension tests varying in their processing demands: a WJPC-Low group (Woodcock-Johnson Passage Comprehension test; \(n = 27\)), a CBM-Maze-Low group (Curriculum-based Measurement-Maze Test; \(n = 18\)), a Recall-Low group (\(n = 19\)), and a control group exhibiting no deficits (\(n = 30\)). All groups of poor readers performed at least one standard deviation below the average age group mean on the respective test used for their identification. The four groups were identified in Grade 2, and they were compared retrospectively in Grade 1 and Kindergarten on a set of cognitive and linguistic measures. The effects of verbal and nonverbal ability, age, and parental education were controlled among the groups. Results showed that the CBM-Maze-Low group exhibited relatively low performance on most linguistic component skills such as RAN, phonological ability, word reading fluency and accuracy, across all three time points. The WJPC-Low and the Recall-Low groups, in contrast, consisted of readers who performed relatively low on word reading fluency and phonological measures only, in Grade 1 and 2 but not in Kindergarten. There were no differences in any of the cognitive measures. Implications of the findings for the use of reading comprehension tests as diagnostic tools are discussed.

\*Corresponding author: Panayiota Kendeou, University of Minnesota, Department of Educational Psychology, 56 East River Rd, Minneapolis, MN 55455, USA. E-mail : kend0040@umn.edu

Authors’ Note. This research was supported in part by European Union–University of Cyprus Grants for Applied Research Projects for Cyprus (No. 8037-16013) and in part by a Cyprus Research Promotion Foundation grant: ΝΕΑΥΠΙΟΔΟΜΗ/ΣΤΡΑΤΗ/0308/37.

L’année psychologique/Topics in Cognitive Psychology, 2014, 114, 725-752
Épreuves d’évaluation de la compréhension en lecture et lecteurs en difficulté :
différents profils émergent des contraintes de traitement associées aux tests

RÉSUMÉ
Cette recherche étudie différents profils de lecteurs faibles grâce au suivi longitudinal d’un groupe initial de 213 enfants, de la grande section de maternelle jusqu’au CE1. Les performances à trois épreuves de compréhension en lecture dont les contraintes de traitement diffèrent ont permis de distinguer quatre groupes de lecteurs: un groupe de faibles lecteurs au test WJPC (Woodcock-Johnson Passage Comprehension test; n = 27), un groupe de faible lecteurs au CBM-Maze test (Curriculum-based Measurement-Maze Test; n = 18), un groupe de faibles lecteurs à une épreuve de rappel (n = 19), et un groupe témoin ne présentant aucun déficit sur les 3 épreuves (n = 30). Dans chaque groupe de lecteurs faibles, les enfants ont obtenu une performance inférieure à au moins un écart-type par rapport à la moyenne de leur groupe d’âge, à l’épreuve qui a servi à leur identification. Les quatre groupes ont été identifiés en CE1, et leurs performances à un ensemble d’habiletés cognitives et linguistiques estimées au cours préparatoire et en grande section de maternelle ont été comparées rétrospectivement. Les effets des capacités verbales et non verbales, de l’âge et du niveau d’éducation des parents ont été contrôlés. Les résultats montrent que le groupe faible au test CBM-Maze a aussi des performances relativement faibles à la plupart des habiletés linguistiques élémentaires prédictives de la lecture, telles que la dénomination rapide, la conscience phonologique et l’exactitude et la fluence d’identification des mots au trois temps de mesure. Au contraire, les groupes faibles au test WJPC et au test de rappel, sont composés de lecteurs ayant des performances relativement faibles en conscience phonologique et en fluence d’identification de mots seulement, au CP et au CE1 mais pas en grande section. Aucune différence dans les performances cognitives générales n’a été observée. L’implication de ces résultats quant à l’utilisation des épreuves de compréhension en lecture à des fins diagnostique est discutée.

Over 30 years of systematic research in reading comprehension have resulted in a number of models and theories about what reading comprehension is and how it should be best developed, assessed, and remediated in the context of reading difficulties. This research has converged to the idea that reading comprehension is a multidimensional construct (Davis, 1944; van den Broek et al., 2005). This conclusion is supported by recent initiatives in the field that explored the relative contribution of component skills on well-known reading comprehension tests (Bowyer-Crane & Snowling, 2005; Cutting & Scarborough, 2006; Francis et al., 2006; Keenan & Betjemann, 2006; Keenan, Betjemann, & Olson, 2008; Nation & Snowling, 1997). The findings of these studies have revealed that common reading comprehension tests exert different linguistic and processing demands to the reader (Kendeou, Papadopoulos, & Spanoudis, 2012). Thus, the use of different reading comprehension tests for diagnostic purposes could potentially result in the identification of distinct groups of
readers who differ in fundamental ways. Our aim in the present study was to explore this hypothesis by identifying different groups of poor readers on the basis of their performance on different reading comprehension tests. To address this aim we identified groups of poor readers in Grade 2 using three different reading comprehension tests and retrospectively examined their reading profiles across several literacy related measures within each grade level. It was decided to define the groups using the Grade 2 scores because previous research has shown that although deficits in reading comprehension may be present from the early school grades, they may not always be apparent before age 8 (Catts, Adlof, & Ellis-Weismer, 2006) and that if these problems are not resolved until age 8, they tend to hamper attainment in all aspects of spoken and written language functioning (Stothard, Snowling, Bishop, Chipchase & Kaplan, 1998).

1. PROCESSING DEMANDS OF READING COMPREHENSION TESTS

When choosing a reading comprehension test as the diagnostic tool (as opposed to the use of component skills), one needs to consider the specific processing demands posed by such test. Indeed, various and vastly different tests are used to assess reading comprehension. These are ranging, for example, from requiring children to fill in missing words in text, selecting the best fitting word in context, retelling the story, applying the knowledge gathered from the text or identifying the theme (van den Broek et al., 2005). Nevertheless, several of these reading comprehension tests seem to have moderate correlations with each other indicating that they don’t necessarily measure the same ability (Keenan, Betjemann, & Olson, 2008). Recent findings show that different reading comprehension tests exert different cognitive processing demands (Kendeou et al., 2012). It is likely, therefore, that the use of different reading comprehension tests as diagnostic tools results in different subtypes of poor readers.

Three tests were used in the present study to assess reading comprehension skills and to define different groups of poor readers: The Woodcock-Johnson Passage Comprehension (WJPC; Woodcock, McGrew, & Mather, 2001), the Curriculum Based Measurement-Maze test (CBM-Maze; Deno, 1985), and a recall test based on Causal Network Theory (van den Broek, 1990).

Previous research has shown that performance on Woodcock-Johnson Passage Comprehension test (WJPC) is strongly depended on phonological and spelling skills (Mehta, Foorman, Branum-Martin, & Taylor, 2005), as
well as decoding (Francis et al., 2006; Keenan et al., 2008). More recently, Kendeou et al. (2012) suggested that performance on the WJPC test is also driven by working memory and orthographic processing skills, with the latter being defined as the ability to recognize the quality of orthographic codes, the speed of accessing those codes, and knowledge of both whole word and subword units. Consequently, children performing poorly on the WJPC tests are likely to exhibit weaknesses in working memory, orthographic processing skills, or both.

Performance on the CBM-Maze test has been shown to be related primarily to efficient word reading (Deno, Maruyama, Espin, & Cohen, 1989; Fuchs et al., 2001; Stahl & Hiebert, 2006) and to skills that support the construction of mental representations of the text during reading, such as inference making (Tolar et al., 2012). With processing speed being an integral component of this test, reading fluency also emerged as a significant predictor of performance on the CBM-Maze test (Kendeou & Papadopoulos, 2012). This means that children who perform poorly on the CBM-Maze are very likely to exhibit weaknesses in fluency tasks.

Finally, it has been reported that performance on the recall test is predicted by working memory (Kellogg, 2007), decoding, and language comprehension skills (Keenan et al., 2008). Given the important role of working memory in phonological and spelling abilities, phonological memory and orthographic processing skills were also among the significant predictors of performance on a recall test (Kendeou et al., 2012). Thus, children who perform poorly on the Recall test are likely to exhibit weaknesses in phonological, working memory, and orthographic processing tasks.

2. CHARACTERISTICS OF POOR READERS

Considering the number of skills associated with reading comprehension, it is not surprising that children with reading comprehension difficulties show impairments on a wide range of language or cognitive tasks tapping working memory (Oakhill, Cain, & Yuill, 1998; Swanson & Berninger, 1995), verbal short-term memory (Cain et al., 2004), reading fluency (Cain & Oakhill, 2006; Juel, 1988; Perfetti, 1985), phonological (Shankweiler, 1989; however see Cain, Oakhill, & Bryant, 2000, for contradictory findings), and orthographic processing skills (Kendeou et al., 2012). In addition, rapid automatized naming (RAN) has been found to distinguish typically developing readers from poor readers during childhood (e.g., Papadopoulos, Georgiou, & Kendeou, 2009; Wolf, Bally, & Morris, 1986).
and in adulthood (Birch & Chase, 2004; Korhonen, 1995) with the relationship between RAN and reading comprehension being influenced by verbal working memory (e.g., Leong, Tse, Loh, & Hau, 2008). Letter identification is also an important indicator, as without being able to identify letters fast and accurately and understand that the letters in words are related to phonemes, children have difficulty becoming proficient decoders, which negatively impacts their reading comprehension (Fletcher et al., 1998). Finally, executive function related skills such as planning and attention have been found to account for additional variance in reading comprehension performance after controlling for individual differences in more proximal skills necessary for reading comprehension including basic decoding skills, reading fluency, and vocabulary (Sesma, Mahone, Levine, Eason & Cutting, 2009).

Importantly, increasing evidence points to a wide range of individual differences among children with reading comprehension difficulties (e.g., Nation, Clarke, Marshall, and Durand, 2004; Nation, Clarke, & Snowling, 2002; Yuill & Oakhill, 1991). This means that different patterns of strengths and weaknesses in poor readers are identified across a range of measures important for text comprehension (e.g., Cornoldi, de Beni, & Pazzaglia, 1996). In the context of the present study, we were particularly interested in two sets of processes that are considered central to reading comprehension in early elementary years which, in turn, are described as unavoidable sources of comprehension difficulties; these are lexical processes and working memory resources (Perfetti, 1985). For this reason, we retrospectively examined (in Grade 1 and Kindergarten) the reading profiles of poor readers identified in Grade 2 on a set of skills that together account for these essential sources of comprehension difficulties in the early years. This set of skills is described next.

3. THE PRESENT STUDY

Our brief review of the literature strengthens the hypothesis that using different reading comprehension tests as diagnostic tools can result in the identification of different subtypes of poor readers with different fundamental weaknesses. To address this issue, we examined retrospectively the reading profiles across several literacy related measures of a relatively large number of children age 8 years old, aiming to provide answers about the developmental profiles of poor readers in the early years of schooling. In doing so, we explored the use of reading comprehension tests as diagnostic
tools taking into account the specific processing demands of each test. Specifically, we hypothesized that defining different groups of poor readers on the basis of their performance on different reading comprehension tests would result in distinct groups of poor readers with different profiles.

Further, we explored the differences of the poor reading comprehension groups that emerged on a large range of linguistic and cognitive skills. With regard to linguistic skills, we included measures of letter knowledge, word decoding, phonological processing, RAN, orthographic processing, and spelling. With regard to cognitive skills we included measures of planning, attention, simultaneous, and successive processing skills following the PASS (Planning, Attention, Simultaneous, and Successive processing) theory of intelligence (Das, Naglieri, & Kirby, 1994). The PASS theory proposes that cognition is organized in three systems and four processes (e.g., Das et al., 1994; Naglieri & Das, 2005; Papadopoulos, Parrila, & Kirby, 2015, for further information on PASS theory). The first system is the Planning system, which involves executive functions responsible for regulating and programming behaviour, selecting and constructing strategies, and monitoring performance. The second system is the Attention system which comprises of basic behaviours such as complex attentional behaviour involved in discrimination learning and selective attention, which in turn, allows the allocation of resources and effort. The third system is the Information processing system which employs Simultaneous and Successive processing, to encode, transform, and retain information. Within this theory, successive processing predicts reading through the effects of phonological awareness, and simultaneous processing predicts reading through the effects of orthographic knowledge (Das, Parrila, & Papadopoulos, 2000; Papadopoulos, 2001; Wang, Georgiou, & Das, 2012). Successive and simultaneous processing demand both processing and storage of information and they have been found to be good direct (particularly successive processing; Georgiou, Das, & Hayward, 2008) or indirect predictors of reading comprehension via phonological and orthographic processing (for successive and simultaneous processing, respectively; see Kendeou, P., Papadopoulos, T. C., & Spanoudis, G. (2015).

3.1. Method
3.1.1. Participants
The original group consisted of 220 children in Cyprus coming from three different districts and 30 urban and rural schools that are typical in Cyprus. Schools were randomly chosen among those that traditionally collaborate with the University of Cyprus for research and training purposes. The children, also randomly selected, were native Greek speakers with no reported history of speech, language, or hearing
Reading comprehension tests and poor readers

difficulties. From this sample, four groups were formed on the basis of a stepwise group selection process: a WJPC-Low group, a CBM-Maze-Low group, a Recall-Low group and a no-deficit group that served as a control.

Step I for group selection: First, we examined the scores on the three reading comprehension tests in Grade 2, namely the Woodcock-Johnson Passage Comprehension test (WJPC), the Curriculum-Based Measurement test (CBM-Maze test) and the Recall test. Children scoring at least one standard deviation\(^1\) below the average age group mean on the WJPC test were included in the WJPC-Low group. Similarly, those scoring at least one standard deviation below the average age group mean on the CBM-Maze test were included in the CBM-Maze-Low group and those children with performance falling at least one standard deviation below the average age group on the Recall test were included in the Recall-Low group. The strict cut-off criterion of the 1 SD (\(\leq 85\) in standard scores or \(\leq 16\) in percentile ranks) was used deliberately and ensured a reasonable degree of separation between the “low” groups and the control. Those few children exhibiting deficits in more than one reading comprehension test (\(n = 7\)) were excluded from further analysis. As a result, those participants who were assigned to a group because of their low score in a specific reading comprehension test, did not exhibit difficulties in any of the other two tests. All other participants (\(n = 149\)) scoring above a standard score of 85 in all three reading comprehension measures formed the initial control group. To assume equal variances across the groups, a smaller number of these participants (\(n = 30\)) were randomly selected to form the final control group.

Step II for group selection: Second, to ensure that reading comprehension difficulties are not confounded with intelligence deficits or demographic variables, groups were compared on verbal (Similarities and Vocabulary, Wechsler Intelligence Scale for Children–Third Edition, Wechsler, 1992; Greek adaptation by Georgas, Paraskevopoulos, Bezevegis, & Giannitsas, 1997) and non-verbal (Matrices, Das-Naglieri Cognitive Assessment System, Naglieri & Das, 1997; Greek adaptation by Papadopoulos, Georgiou, Kendeou, & Spanoudis, 2008) ability, parental education, and age. The findings showed that groups did not differ in any of these selection variables: verbal and nonverbal-ability, Wilks \(\lambda, F(9, 214.32) = 0.60, ns\); parental education, \(\chi^2(21,94) = 19.87, ns\) or age \(F(3, 90) = 1.26, ns\). The final groups were as follows: (a) WJPC-Low group (\(n = 27, 17\) girls and 10 boys), (b) CBM-Maze-Low group (\(n = 18, 9\) girls and 9 boys); (c) Recall-Low group (\(n = 19, 15\) girls and 4 boys); and (d) control group exhibiting no deficits (\(n = 30; 10\) girls and 20 boys). The mean age of this final set of children (a total \(n = 94\) cases out of the original 220) in the initial assessment (kindergarten) was 5 years 8 months (SD = 0.30 years, minimum = 5.2, maximum = 6.4), in the second assessment (in Grade 1), 6 years 6 months (SD = 0.30 years, minimum = 6.1, maximum = 7.2), and in the final assessment (in Grade 2) 7 years 6 months (SD = 0.40 years, minimum = 7.0, maximum = 8.2). Participants’ characteristics and group scores are summarized in Table 1. All groups performed within normal range in the word-reading tasks in Grade 2 (see Table 2). In addition, none of

\(^{1}\)According to the Diagnostic and Statistical Manual of Mental Disorders, 5th edition (DSM-5; APA, 2013) the precise scores for the diagnosis of specific learning disorders, such as reading difficulties, may vary depending on the particular standardized tests used. Thus, it is recommended that on the basis of clinical judgment, a more lenient threshold to be used, with an onset of 1.0 SD below the population mean for age.

L’année psychologique/Topics in Cognitive Psychology, 2014, 114, 725-752
### Table 1. Data on the Demographic and Ability Variables for WJPC-Low, CBM-Maze-Low, Recall-Low, and Control Groups

<table>
<thead>
<tr>
<th>Variables</th>
<th>Groups</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>WJPC-Low (n = 27)</td>
</tr>
<tr>
<td>Age Mean (SD)</td>
<td>7.65 (0.35)</td>
</tr>
<tr>
<td>Age Range</td>
<td>0.86</td>
</tr>
<tr>
<td>Gender Females</td>
<td>17 [62.9%]</td>
</tr>
<tr>
<td>Gender Males</td>
<td>10 [37.1%]</td>
</tr>
<tr>
<td>Parental Education Level</td>
<td></td>
</tr>
<tr>
<td>Less than HS</td>
<td>3 [11.1%]</td>
</tr>
<tr>
<td>HS Graduate</td>
<td>12 [44.4%]</td>
</tr>
<tr>
<td>Some College</td>
<td>7 [25.9%]</td>
</tr>
<tr>
<td>College Graduate</td>
<td>5 [18.5%]</td>
</tr>
<tr>
<td>Non-Verbal Ability</td>
<td></td>
</tr>
<tr>
<td>CAS Matrices</td>
<td>98.73 (14.37)</td>
</tr>
<tr>
<td>Verbal Ability</td>
<td></td>
</tr>
<tr>
<td>Similarities</td>
<td>96.88 (14.48)</td>
</tr>
<tr>
<td>Vocabulary</td>
<td>101.27 (12.04)</td>
</tr>
<tr>
<td>Reading Comprehension</td>
<td></td>
</tr>
<tr>
<td>WJPC</td>
<td>77.72 (13.17)</td>
</tr>
<tr>
<td>CBM-Maze</td>
<td>98.26 (12.34)</td>
</tr>
<tr>
<td>Recall</td>
<td>101.02 (9.92)</td>
</tr>
</tbody>
</table>

Note: WJPC = Woodcock-Johnson Passage Comprehension test; CBM-Maze = Curriculum-Based Measurement test; Values in parentheses are SDs; values in brackets are percentages; the reported ability and reading comprehension scores are standardized scores; Non-Verbal and Verbal Ability scores were available only for Grade 1.
the children in the groups of poor readers had been identified earlier as reading
disabled. All children were classified as normally achieving readers before Grade 2.

3.1.2. Measures

3.1.2.1. Reading Measures (used as selection criteria)

Reading Comprehension Measures. Three tests were administered to the
participants to assess reading comprehension skills. The Woodcock-Johnson
Passage Comprehension (WJPC; Woodcock, McGrew, & Mather, 2001), the
Curriculum Based Measurement-Maze test (CBM-Maze; Deno, 1985), and a
recall test based on Causal Network Theory (van den Broek, 1990). These
tests were adapted and initially used in Greek by Papadopoulos, Georgiou, and
Kendeou (2009). Following the study objectives, reading comprehension tests were
administered only in Grade 2.

W-J Passage Comprehension. In this test, participants were required to read
short passages silently (usually two to three lines long) and provide a missing word
(represented by a blank line). The task contained 68 items. A participant’s total
score was the number of correctly filled blanks. Cronbach’s alpha in Grade 2 in the
present study was .72.

CBM-Maze test. In this test, participants were required to read passages and
choose the correct word among three options. Participants encountered three
options every seventh word in each passage. In the present study, participants read
a total of three passages, one at a time, in a booklet form with 155, 176, and 183
words, respectively. These passages were similar to narrative texts that participants
might be exposed to in their own reading or in school. Participants were given 1
minute to read as much of each passage as possible and, while reading, circle the ap-
propriate word in context (for task properties see Pierce, McMaster, & Deno, 2010).
A participants’ score consisted of the average number of correct words chosen
across the three texts minus the average number of incorrect words chosen (follow-
ing Fuchs et al., 2001). Cronbach’s alpha in Grade 2 in the present study was .82.

Recall test. In this test, participants were asked to read and recall an age
appropriate narrative that had a standard but complex story structure in which
the protagonist made several attempts to achieve his desired goal (story length
was 177 words). After reading, the experimenter asked each participant to “tell
everything you remember from the story from the beginning.” If a participant
did not recall any narrative events spontaneously, the experimenter asked a more
specific question, “What happened at the beginning of the story?” Participants were
prompted to continue to recall the story (i.e., “What else do you remember?”),
until they indicated that they could not recall anything else. A participant’s score
consisted of the total number of events recalled, central and non-central to the
story causal structure. Central events were defined as those that had three or more
causal connections to other events in the text. To identify central and non-central
events, causal relations between all events in the story were identified according to
principles of causality (van den Broek, 1990). Cronbach’s alpha for this task was calculated between the recall of events that were central and non-central to the causal structure, and in Grade 2 it was equal to .65. Two independent raters coded 18% of the recall protocols in common and inter-rated reliability was .82.

Word Reading Measures. Three standardized measures from the Early Reading Skills Assessment Battery (ERS-AB; Papadopoulos, Spanoudis, & Kendeou, 2008), a real word, a nonword, and a silent word reading task were used to assess participants’ word reading ability. Papadopoulos et al. (2008) reported Cronbach’s alpha for the real word reading task to be .88, .97, and .81 and for the nonword reading task .92, .92, and .70 in Kindergarten, Grade 1, and Grade 2, respectively. Cronbach’s alpha for the silent word reading task was .72 in Grade 1 and .83 in Grade 2. In the read aloud tasks, participants were asked to read the entire list of words as accurately and quickly as possible. Both, the accuracy score (the total number of words read correctly) and reading fluency score (the number of words read correctly within 60 seconds), were recorded for each participant.

Word Identification (WID). This test consisted of 80 words forming a $2 \times 2$ factorial design in terms of frequency (high/low), orthographic regularity (regular/exception), and length (bisyllable/trisyllable). The stimulus words were mainly nouns with a few adjectives and verbs.

Word Attack (WAT). This test consisted of 45 pronounceable nonwords that were derived from real words after changing two or three letters (either by substituting them or using them backwards). The task started with bisyllabic words and ended with five-syllabic words.

Silent Word Reading: Word Chains (WC). This task required children to scan words presented as a continuous line of print without inter-word spaces (e.g., “under/word/leave”). They were given one minute to identify the words in each row by drawing a slash to indicate where the spaces should be (e.g., “under/word/leave”). The test included a total of 15 rows of words of increasing length. The first two rows consisted of two words put together whereas the last three items consisted of seven words put together. The score was the number of correctly placed slashes.

3.1.2.2. Verbal Ability Measures

WISC-III: Similarities (SIM) and Vocabulary (VOC). The Vocabulary and Similarities subtests from the Wechsler Intelligence Scale for Children-III (WISC-III-R; Wechsler, 1992; Greek adaptation: Georgas et al., 1997) were used to assess verbal ability. In the Vocabulary subtest, participants were asked to provide verbal definitions for words presented orally by the examiner (30 items). Participants’ answers reflecting a relevant general definition earned 2 points, whereas responses reflecting only one or more common properties of an item earned 1 point. The task was discontinued after four consecutive mistakes. The total score was the participants’ final score. Georgas et al. (1997) reported Cronbach’s alpha reliability coefficient to be .68, for Grade 1. The Similarities subtest was used
to measure verbal abstract reasoning and conceptualization abilities. Participants were asked to describe how two things were alike (e.g., how a fork and a knife are alike?; 19 items). Participants’ answers reflecting a relevant general verbal analogy earned 1 point, whereas responses not reflecting the shared characteristic between two concepts earned no points. The task was discontinued after four consecutive mistakes. The total score based on the sum of 1-point responses was the participants’ final score. The reported Cronbach’s alpha for this task in Grade 1 is .65.

3.1.2.3. Non-Verbal Ability Measure

CAS Matrices (MAT). The Matrices subtest from the Cognitive Assessment System (CAS; Naglieri & Das, 1997) was administered as a measure of non-verbal ability. This was a 33-item multiple choice test that utilized shapes and geometric designs that were interrelated through spatial or logical organization. Participants were required to decode the relationships among the parts of the item and choose the best of six options. Each progressive matrix item was scored as correct or incorrect. The raw score was the total number of items correctly answered. Papadopoulos, Georgiou, Kendeou et al. (2008) reported Cronbach’s alpha reliability coefficient to be .72 in Grade 1.

3.1.2.4. Linguistic Measures

Phonological Ability. Participants’ phonological skills were assessed using the Phoneme Elision and Phoneme Blending tasks from a standardized battery of phonological tasks (Papadopoulos, Kendeou, & Spanoudis, 2012; Papadopoulos, Spanoudis, & Kendeou, 2009). Both tasks tapped phonological ability at the phonemic level. Both tasks included 15 testing items and were discontinued after four consecutive failures. In both instances, a participant’s score was the total number of correct responses.

Phoneme Elision (PE). In this task, participants were asked to repeat a word after deleting an identified phoneme. The targeted phonemes were either vowels or consonants and their position varied systematically (from beginning and final to middle phoneme) across items. After deleting the target phoneme, the remaining phonemes formed a word (e.g., say the word /τωρα/; /ora/; now, after deleting the sound /t/; /ωρα/; /ora/; time). Papadopoulos et al. (2012) reported Cronbach’s alpha reliability coefficient to be .93, .93, and .88, for Kindergarten, Grade 1, and Grade 2, respectively.

Phoneme Blending (BL). In this task, audio prompts presented the sounds of two-to-six sound words separately, and participants were asked to orally blend them into a word. Word complexity was progressively more difficult. The first four words consisted of two- to four- phoneme segments that were of CV or CVC structure (e.g., /φως/; /fos/; light). The more difficult items contained more complex phoneme segments such as CCV (e.g., /στόμα/; /stoma/; mouth). Papadopoulos et al. (2012) reported Cronbach’s alpha reliability coefficient to be .91, .90, and .85, for Kindergarten, Grade 1, and Grade 2, respectively.
Rapid Automatized Naming (RAN). This set of tasks was originally developed and used by Papadopoulos, Charalambous, Kanari, & Loizou (2004). All four measures included two tasks (one relatively easy and one more difficult) made up of 20 testing items (5 different stimuli, each repeated four times). The items in each task were presented on a single page, with four lines of 5 items per page. Order of items changed from one line to the other. Participants were asked to name the stimuli as quickly as possible. Participants’ score was the ratio of the total number of items named correctly divided by the total time taken.

RAN-Pictures (RAN-P). This measure was modeled after Wimmer, Mayringer, and Landerl (2000). The words of the first task started with the same single consonant cluster (καπέλο/καρέκλα/κεράτο/κέλιδι/καπέλο/καρέκλα/κεράτο/κέλιδι; hat, chair, cherry, carrot, key) whereas the words of the second task started with different consonant clusters (φράουλα/πλιτίριο/σκιλόσ/σταύρος/μπανάνα; fraoula / plintirio / skilos / stavros / banana). In the present sample, the correlation between successive Grades was .44 from Kindergarten to Grade 1.

RAN-Colors (RAN-C). Five basic and relatively more frequent colors, namely, red, green, yellow, blue, and white were included in the first task. In contrast, the second task was comprised of less frequent and secondary colors such as “pink”, “light blue”, “brown”, “orange”, and “purple”. In the present sample, the correlation between successive Grades was .34 from Kindergarten to Grade 1.

RAN-Digits (RAN-D). The digits from 1 to 5 were included in the first task. The second task was comprised of the digits 6 to 9 and 0 (zero). In the present sample, the correlation between successive Grades was .47 from Grade 1 to Grade 2.

RAN-Letter (RAN-L). The letters included in the first task were only vowels (α, η, ε, ο, ν); and the letters in the second task were only consonants that shared similar characteristics (π, τ, σ, δ, θ). In the present sample, the correlation between successive Grades was .46 from Grade 1 to Grade 2.

3.1.2.5. Letter Knowledge

Letter Identification. The Letter Identification subtest from the Dyslexia Early Screening Test-2 battery (DEST-2; Nicolson & Fawcett, 2004; Greek standardization: Papadopoulos, Georgiou, & Kendeou, 2008) was administered to assess letter knowledge. This test was selected because it provided information about participants’ ability to identify different and relatively frequent letters in Greek (3 vowels and 7 consonants in total). The letters were presented to the participants on a single page with large font size (Arial 96) in lowercase. Participants were asked to provide the name of the letter. Participants’ score was the total number of correct responses. Papadopoulos, Georgiou and Kendeou (2008) reported Cronbach’s alpha reliability coefficient to be .65 and .94, for Kindergarten and Grade 1, respectively.
Orthographic Processing. Two measures were used to assess participants’ orthographic processing skills, namely, orthographic choice (also drawn from ERS-AB) and two-minute spelling. In both measures, participants’ score was the total number of correct responses.

Orthographic Choice (OC). This task consisted of 20 items that were constructed in a way that phonological transcription alone did not reliably result in identifying the one orthographically correct word among the three words included in each item (e.g., /αρέσι/; like). Participants had to use their knowledge of the orthographic patterns for the given words in order to identify the one that was both phonologically and orthographically correct. Cronbach’s alpha reliability coefficient in our sample was .67 in Grade 1 and .80 in Grade 2.

Two-Minute Spelling (TMS). This task was taken from the Dyslexia Screening Test-Junior (DST-J; Fawcett & Nicolson, 2004; Greek standardization: Papadopoulos, Georgiou, & Spanoudis, 2008) and was used to assess participants’ spelling skills. This task involves speed of writing as well as accuracy of the spelling. Participants were asked to spell a certain amount of words (up to 32 two- to multi-syllabic words) within two minutes. Cronbach’s alpha for this task in the Greek standardization sample was .80 in Grade 2.

3.1.2.6. Cognitive Measures

Eight measures taken from the Das-Naglieri Cognitive Assessment System (DN-CAS; Naglieri & Das, 1997; Greek standardization: Papadopoulos, Georgiou, Kendeou, & Spanoudis, 2008) were used to assess Planning, Attention, Simultaneous, and Successive processing. All Cronbach’s alpha reported below for the CAS tests are based on the Greek standardization.

3.1.2.7. Planning

Planned Connections. The task required the participant to develop some effective way to connect sequential stimuli (numbers 1-2-3-4-5-...), which were quasi-randomly distributed on a page. In this study, the task consisted of six items: the first two items required the child to join in sequence the numbers from one to five, the next two had numbers up to ten, the fifth one had 15 numbers, and the last item, up to number 25. The participant used a red pencil to join the numbers and his or her Planned Connection score was the combined time to complete items 1 to 6. Cronbach’s alpha reliability coefficient was .77, .70, and .76, for Kindergarten, Grade 1, and Grade 2, respectively.

Matching Numbers (MN). This task consisted of four pages each consisting of eight rows of numbers with six numbers per row. Children were instructed to underline the two numbers in each row that were the same. Numbers increased in length across the four pages from one digit to seven digits with four rows for each
digit length. Each item had a time limit. Children 5 to 7 years are administered Items 1 and 2. The subtest score was based on the combination of time and number correct (accuracy score) for each page. Accuracy scores were summed and used as a measure of the child’s efficiency. Cronbach’s alpha reliability coefficient was .67, .73, and .67, for Kindergarten, Grade 1, and Grade 2, respectively.

3.1.2.8. Attention

Expressive Attention. This task is based on the Stroop task. The version used in the first two years of this study was composed of three pages, all of them containing animals. Participants were shown animals that were either “small” (a butterfly, a mouse, a bird, and a frog) or “big” (an elephant, a whale, a horse, and a dinosaur). In Item 1, all of the pictures were of the same physical size (i.e., all animals were present in a relatively big size, regardless of their relative actual size); in Item 2, the size of the pictorial representation was in accordance with actual size differences (i.e., pictures of small animals were smaller than pictures of large animals); and in Item 3, the pictorial presentations of the animals did not follow their actual size, but instead, small and big animals could be presented either as small or big. The participants were required to label all pictures in the item as representing either big or small animals. The participant’s Expressive Attention score was the ratio score of the item 3 completion time divided by the number of correct responses in this item. In the Grade 2 testing, following the standard DN-CAS administration procedures (Naglieri & Das, 1997), the original Stroop task with color-names was used as the Expressive Attention task. Again, the ratio score, which derived from the division of time by number of correct responses in item 6 was used as a participant’s Expressive Attention score. Cronbach’s alpha reliability coefficient was .86, .85, and .72, for Kindergarten, Grade 1, and Grade 2, respectively.

Receptive Attention. In this task, in Kindergarten and Grade 1, participants were given four sheets consisting of 50 picture pairs each (trees, fruits, flowers, birds, houses, or human faces) arranged in a matrix form. In the first two items, the participants’ task was to underline only those pairs of pictures that were visually alike (picture matching). Alternatively, in the last two items, the participants were instructed to underline those pairs that belonged to the same taxonomic category (name matching). A participant’s Receptive Attention score was the combined time to complete items 3 and 4 divided by the total number of correct responses in these items. In Grade 2, this task used letters instead of images. Cronbach’s alpha reliability coefficient was .85, .82, and .78, for Kindergarten, Grade 1, and Grade 2, respectively.

3.1.2.9. Simultaneous Processing

Figure Memory (FM). This task consisted of 27 geometric designs, such as a triangle or a square, that were presented to the participant one at a time for a period of five seconds each. Following the presentation of a particular target design
participants were given a more complex design in which the target design was embedded. Participants were then asked to outline the original target. The task was discontinued after 4 consecutive failures. A participant’s score was the total number of items correctly reproduced. Cronbach’s alpha reliability coefficient was .76, .77, and .83, for Kindergarten, Grade 1, and Grade 2, respectively.

**Verbal-Spatial Relations (VS).** This task was composed of 27 items that required the comprehension of logical and grammatical descriptions of spatial relationships. Participants were shown items containing six drawings and a printed question at the bottom of each page. The items involved both objects and shapes that were arranged in a specific spatial manner. For example, the item “Which picture shows a circle to the left of a cross under a triangle above a square?” included six drawings with various arrangements of geometric figures, only one of which matched the description. The examiner read the question aloud and the child was required to select the option that matched the verbal description. Participants had to indicate their answer within the 30-second time limit to receive credit. A participant’s score was the total number of items correctly answered. Cronbach’s alpha reliability coefficient was .72, .73, and .75, for Kindergarten, Grade 1, and Grade 2, respectively.

**3.1.2.10. Successive Processing**

**Sentence Repetition and Questions (SRQ).** This was a serial memory task for words and repeating nonsense sentences, engaging both processing and storage of information. The task consisted of two parts. Participants had to first repeat and then answer questions about nonsensical sentences in which the content words were replaced by color words (e.g. “The yellow greened the blue”). Thus, participants could use syntactic cues but no semantic cues to remember the sentences or to answer the questions. A participant’s score was the number of correctly reproduced sentences plus the number of correctly answered questions. Cronbach’s alpha reliability coefficient was .73, .75, and .77, for Kindergarten, Grade 1, and Grade 2, respectively.

**Speech Rate (SpR).** This task required participants to repeat a high imagery, single-, and double-syllable word series 10 times in order. Participants were timed to determine how long it takes to repeat the series correctly. There were 8 items. Examiners began timing when participants said the first word in the series and stopped timing when participants finished repeating the last word in the tenth repetition. A participant’s score was the total time in seconds for all items. Cronbach’s alpha reliability coefficient was .92, .94, and .92, for Kindergarten, Grade 1, and Grade 2, respectively.

**3.1.3. Procedure**

In all three assessments, participants were tested individually in a session lasting approximately 60 minutes, between February and April each year. All participants
were available and tested in all three years of the study. The presentation of the tasks was held constant across years. With some exceptions, the same set of tasks was administered to all participants across measurement points. Specifically, Letter Knowledge was administered only in Kindergarten and Grade 1, Orthographic Choice and Silent Word Reading in Grade 1 and Grade 2, and Two-Minute Spelling only in Grade 2. RAN-Digits and RAN-Letters were used in all analyses performed in Grades 1 and 2, and RAN-Colors and RAN-Pictures instead were used in Kindergarten. All testing took place during school hours in a private room in the participants’ respective schools. Examiners were trained graduate research assistants enrolled in educational psychology courses, blind to grouping of children. None of the participants in the deficit groups received systematic intervention in their respective schools over the course of the study. Written permission from schools and parents was obtained prior to testing.

3.2. Results

Before performing any analyses we computed composite scores expressed in standardized T-score units (with a mean of 100 and a standard deviation of 15) averaged across measures for each of those set of skills for which scores from two measures at two or all three time points were available. These calculations followed the procedures outlined by Naglieri and Das (1997) and were possible given the relatively medium to high correlations between the different measures. Specifically, the phonological ability composite score was calculated using the phoneme elision and phoneme blending tasks ($r$ ranged from .80 in Kindergarten to .76 in Grade 1). RAN composite score was calculated using the non-alphanumeric and alphanumeric RAN tasks ($r$ ranged from .63 in Grade 2 to .47 in Kindergarten). Successive processing composite score was calculated using the sentence questions repetition and speech rate tasks ($r$ ranged from -.49 in Grade 1 to -.43 in Grade 2). Simultaneous processing composite score was calculated using verbal-spatial relations and figure memory tasks ($r$ ranged from .25 in Grade 1 to .22 in Kindergarten). Attention composite score was calculated using expressive and receptive attention tasks ($r$ ranged from .36 in Kindergarten to .29 in Grade 1). Planning composite score was calculated using planned connections and matching numbers tasks ($r$ ranged from -.49 in Kindergarten to -.41 in Grade 2). Similarly, reading fluency was calculated using word identification and word attack (60-second performance; $r$ ranged from .95 in Kindergarten to .64 in Grade 2) and reading accuracy was calculated using the total number of words read in both measures ($r$ ranged from .94 in Kindergarten to .37 in Grade 2). Table 2 shows the group scores on all composite and single measures (T-scores are presented) at all three measurements points.
Table 2. Descriptive Statistics for Deficit and Control Groups on Reading, Linguistic, and Cognitive Measures in Grade 2

<table>
<thead>
<tr>
<th>Variables</th>
<th>WJPC-Low (n = 27)</th>
<th>CBM-Maze-Low (n = 18)</th>
<th>Recall-Low (n = 19)</th>
<th>Control (n = 30)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Word Fluency</td>
<td>97.16 (14.07)</td>
<td>90.10 (14.51)</td>
<td>99.43 (7.11)</td>
<td>106.10 (12.76)</td>
</tr>
<tr>
<td>Word Accuracy</td>
<td>98.59 (14.08)</td>
<td>94.85 (14.11)</td>
<td>103.45 (8.70)</td>
<td>104.82 (12.01)</td>
</tr>
<tr>
<td>Silent Word Reading</td>
<td>97.51 (19.02)</td>
<td>97.21 (14.13)</td>
<td>100.44 (11.35)</td>
<td>104.17 (13.69)</td>
</tr>
<tr>
<td>Phonological Processing</td>
<td>96.92 (14.19)</td>
<td>96.74 (13.37)</td>
<td>102.95 (8.27)</td>
<td>107.54 (6.00)</td>
</tr>
<tr>
<td>RAN</td>
<td>95.56 (13.98)</td>
<td>94.76 (9.49)</td>
<td>95.30 (11.48)</td>
<td>103.41 (11.46)</td>
</tr>
<tr>
<td>2-Minute Spelling</td>
<td>101.84 (15.11)</td>
<td>96.18 (17.45)</td>
<td>98.97 (13.05)</td>
<td>97.03 (14.31)</td>
</tr>
<tr>
<td>Orthographic Processing</td>
<td>98.45 (14.49)</td>
<td>96.97 (11.10)</td>
<td>104.50 (8.26)</td>
<td>104.63 (8.93)</td>
</tr>
<tr>
<td>Planning</td>
<td>99.72 (9.04)</td>
<td>103.41 (6.31)</td>
<td>102.34 (7.49)</td>
<td>100.17 (8.24)</td>
</tr>
<tr>
<td>Attention</td>
<td>99.35 (12.34)</td>
<td>97.54 (9.85)</td>
<td>102.38 (10.85)</td>
<td>100.14 (10.59)</td>
</tr>
<tr>
<td>Successive Processing</td>
<td>99.94 (7.21)</td>
<td>101.41 (7.57)</td>
<td>99.05 (8.04)</td>
<td>102.22 (8.41)</td>
</tr>
<tr>
<td>Simultaneous Processing</td>
<td>98.69 (9.81)</td>
<td>94.37 (12.84)</td>
<td>96.53 (12.70)</td>
<td>104.79 (10.36)</td>
</tr>
</tbody>
</table>

Note: Values in parentheses are SDs; the reported scores are standardized scores.

3.2.1. Comparison of WJPC-Low, CBM-Maze-Low, Recall-Low and Control Groups

Grade 2. In Grade 2, a series of ANOVAs were performed with group as a fixed factor and phonological abilities and RAN as the dependent variables. The main group effect was significant only for phonological abilities, $F(3, 90) = 5.98$, $p < .05$, $\eta^2 = .17$. Post-hoc comparisons using Bonferroni adjustment showed that both the WLPC-Low ($p = .002$) and the CBM-Maze-Low ($p = .007$) groups scored significantly lower than the
Control group (Table 2). Next, a series of MANOVAs were performed with group as a fixed factor and word reading (accuracy and fluency), orthographic processing, and cognitive measures as the dependent variables (4 groups × 4 tasks). The analysis showed a group main effect for word reading measures only, Wilks’ $\lambda = .80$, $F(9, 214.32) = 2.27$, $p < .05$. Subsequent univariate analyses demonstrated that the main effect of group was significant for two out of the four measures, specifically word fluency $F(3, 90) = 6.26$, $p < .05$, $\eta^2 = .10$ and word accuracy $F(3, 90) = 3.46$, $p < .05$, $\eta^2 = .04$. In turn, post-hoc tests using Bonferroni adjustment, showed that the CBM-Maze-Low group scored significantly lower than the Control group in both word fluency ($p = .001$) and word accuracy ($p = .029$).

**Grade 1.** In Grade 1, a series of ANOVAs were performed with group as a fixed factor and phonological, RAN, orthographic processing, and letter knowledge as the dependent variables. The main group effect was significant for phonological abilities, $F(3, 90) = 7.65$, $p < .05$, $\eta^2 = .20$ and RAN, $F(3, 90) = 4.56$, $p < .05$, $\eta^2 = .13$. Post-hoc comparisons using Bonferroni adjustment showed that for the phonological measures both the WLPC-Low ($p = .013$) and the CBM-Maze-Low ($p = .001$) groups scored significantly lower than the Control group (Table 3). For naming speed, the CBM-Maze-Low group performed significantly lower than the Control ($p = .008$). Next, a series of MANOVAs were performed with group as a fixed factor and word reading and cognitive measures as the dependent variables. The analysis showed a group main effect for word reading measures only (4 groups x 3 tasks), Wilks’ $\lambda = .78$, $F(9, 214.32) = 2.51$, $p < .05$. Subsequent univariate analyses demonstrated that the main effect of group was significant for two out of the three measures, specifically word fluency $F(3, 90) = 5.97$, $p < .05$, $\eta^2 = .17$ and word accuracy $F(3, 90) = 3.82$, $p < .05$, $\eta^2 = .12$. In turn, post-hoc tests using Bonferroni adjustment, showed that all three deficit groups scored significantly lower than the Control group in word fluency ($p = .004$ for WJPC-Low; $p = .027$ for CBM-Maze-Low; $p = .007$ for Recall-Low). In word accuracy the CBM-Maze-Low group scored significantly lower than the Control group ($p = .022$).

**Kindergarten.** In Kindergarten, a series of ANOVAs were performed with group as a fixed factor and phonological, RAN, and letter knowledge as the dependent variables. The main group effect was significant only for RAN, $F(3, 90) = 4.99$, $p < .05$, $\eta^2 = .13$ and letter knowledge, $F(3, 90) = 2.86$, $p < .05$, $\eta^2 = .09$. Post-hoc comparisons using Bonferroni adjustment showed that for RAN the CBM-Maze-Low ($p = .003$)
Table 3. Descriptive Statistics for Deficit and Control Groups on Reading, Linguistic, and Cognitive Measures in Grade 1

<table>
<thead>
<tr>
<th>Variables</th>
<th>WJPC-Low (n = 27)</th>
<th>CBM-Maze-Low (n = 18)</th>
<th>Recall-Low (n = 19)</th>
<th>Control (n = 30)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>M (SD)</td>
<td>M (SD)</td>
<td>M (SD)</td>
<td>M (SD)</td>
</tr>
<tr>
<td>Word Fluency</td>
<td>96.89 (11.58)</td>
<td>97.76 (13.52)</td>
<td>96.16 (10.78)</td>
<td>109.31 (15.74)</td>
</tr>
<tr>
<td>Word Accuracy</td>
<td>99.76 (13.71)</td>
<td>94.87 (15.05)</td>
<td>96.39 (12.81)</td>
<td>106.04 (9.36)</td>
</tr>
<tr>
<td>Silent Word Reading</td>
<td>97.88 (14.52)</td>
<td>94.66 (12.37)</td>
<td>100.41 (12.69)</td>
<td>103.56 (15.63)</td>
</tr>
<tr>
<td>Phonological Processing</td>
<td>97.52 (14.08)</td>
<td>91.56 (12.82)</td>
<td>101.72 (11.20)</td>
<td>107.24 (7.88)</td>
</tr>
<tr>
<td>RAN</td>
<td>98.59 (14.72)</td>
<td>94.27 (11.94)</td>
<td>96.96 (7.17)</td>
<td>106.82 (13.87)</td>
</tr>
<tr>
<td>Letter Knowledge</td>
<td>99.70 (16.45)</td>
<td>98.11 (20.15)</td>
<td>98.37 (19.61)</td>
<td>102.58 (1.56)</td>
</tr>
<tr>
<td>Orthographic Processing</td>
<td>99.92 (16.30)</td>
<td>95.62 (17.81)</td>
<td>101.97 (10.74)</td>
<td>105.32 (11.79)</td>
</tr>
<tr>
<td>Planning</td>
<td>98.93 (7.14)</td>
<td>99.34 (10.58)</td>
<td>101.12 (9.76)</td>
<td>98.27 (6.33)</td>
</tr>
<tr>
<td>Attention</td>
<td>101.62 (12.45)</td>
<td>97.84 (9.88)</td>
<td>97.85 (7.73)</td>
<td>101.74 (10.99)</td>
</tr>
<tr>
<td>Successive Processing</td>
<td>100.31 (6.87)</td>
<td>101.21 (8.89)</td>
<td>102.83 (7.41)</td>
<td>101.43 (7.38)</td>
</tr>
<tr>
<td>Simultaneous Processing</td>
<td>98.09 (9.82)</td>
<td>97.19 (13.16)</td>
<td>97.11 (12.61)</td>
<td>102.85 (10.92)</td>
</tr>
</tbody>
</table>

Note: Values in parentheses are SDs; the reported scores are standardized scores.

Group scored significantly lower than the Control group (Table 4). For letter knowledge also the CBM-Maze-Low group performed significantly lower than the Control ($p = .046$). Next, a series of MANOVAs were performed with group as a fixed factor and word reading and cognitive measures as the dependent variables. The analysis showed no group main effects.
Table 4. Descriptive Statistics for Deficit and Control Groups on Reading, Linguistic, and Cognitive Measures in Kindergarten

<table>
<thead>
<tr>
<th>Groups</th>
<th>WJPC-Low (n = 27)</th>
<th>CBM-Maze-Low (n = 18)</th>
<th>Recall-Low (n = 19)</th>
<th>Control (n = 30)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Word Fluency</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>M (SD)</td>
<td>103.24 (19.49)</td>
<td>95.47 (0.0)</td>
<td>99.42 (9.73)</td>
<td>103.75 (20.29)</td>
</tr>
<tr>
<td>Word Accuracy</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>M (SD)</td>
<td>101.74 (14.67)</td>
<td>95.86 (0.0)</td>
<td>98.85 (7.31)</td>
<td>104.32 (25.84)</td>
</tr>
<tr>
<td>Phonological Processing</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>M (SD)</td>
<td>103.72 (18.91)</td>
<td>95.77 (7.35)</td>
<td>97.07 (8.73)</td>
<td>104.89 (19.85)</td>
</tr>
<tr>
<td>RAN</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>M (SD)</td>
<td>97.73 (15.23)</td>
<td>91.06 (13.96)</td>
<td>98.47 (10.12)</td>
<td>105.65 (13.37)</td>
</tr>
<tr>
<td>Letter Knowledge Planning</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>M (SD)</td>
<td>102.07 (12.28)</td>
<td>92.52 (11.61)</td>
<td>96.92 (13.12)</td>
<td>108.06 (28.49)</td>
</tr>
<tr>
<td>Attention</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>M (SD)</td>
<td>102.14 (7.26)</td>
<td>97.35 (8.98)</td>
<td>99.04 (7.31)</td>
<td>102.02 (7.49)</td>
</tr>
<tr>
<td>Successive Processing</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>M (SD)</td>
<td>98.60 (12.31)</td>
<td>94.59 (11.92)</td>
<td>100.77 (7.78)</td>
<td>102.70 (11.98)</td>
</tr>
<tr>
<td>Simultaneous Processing</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>M (SD)</td>
<td>100.28 (8.42)</td>
<td>100.11 (4.57)</td>
<td>100.58 (6.15)</td>
<td>100.01 (7.36)</td>
</tr>
<tr>
<td>Note: Values in parentheses are SDs; the reported scores are standardized scores.</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

4. DISCUSSION

In the present study, we hypothesized that defining different groups of poor readers on the basis of their performance on different reading comprehension tests, would result in distinct groups of poor readers with unique profiles. We identified these groups of poor readers in Grade 2 and retrospectively examined their reading profiles across several linguistic and cognitive measures in Grade 1 and in Kindergarten to explore any “at-risk” signs.

On the basis of our approach and results we reached a better understanding of the diagnostic validity of different reading comprehension tests. Our findings showed that readers, who were identified as poor on the basis of the CBM-Maze test in Grade 2, reliably differed from the control group
on a number of component skills, such as RAN and letter knowledge (in K), phonological (in Grades 1 and 2), and word reading fluency and accuracy (in Grades 1 and 2). The findings also showed that readers, who were identified as poor on the basis of the WJPC test in Grade 2, reliably differed from the control group on a rather small number of component skills, specifically on phonological (in Grades 1 and 2) and word reading fluency (in Grade 1). Finally, readers who were identified as poor on the basis of the Recall test in Grade 2, reliably differed from the control group only on word reading fluency (in Grade 1). Most important, the three distinct groups did not differ from each other in Kindergarten, Grade 1 or Grade 2 in any of the linguistic and cognitive skills measured in this study.

Furthermore, in terms of the profiles of poor readers as a result of each of these reading comprehension tests, the findings showed that the CBM-Maze-Low group tends to be a group of readers with relatively low performance on most linguistic component skills used in the present study, namely RAN, phonological measures, word reading fluency, and word reading accuracy, but only when compared to the typical control group. The WJPC-Low and the Recall-Low groups, in contrast, tend to consist of readers who perform relatively low on word reading fluency and phonological measures only, also when compared to the typical control. Notably, none of the groups of poor readers showed any cognitive deficits compared to their typical counterparts.

Overall the present findings suggest that even though the use of different reading comprehension tests as diagnostic tools can result in the identification of distinct groups of poor readers in Grade 2, this approach provides only partial information about the specific reading comprehension difficulties these readers likely experience at the level of specific linguistic and cognitive component skills. Specifically, any differences observed in specific components skills were only between a deficit group and the typical control group; there were no differences between the three deficit groups in any of the component skills measured, linguistic or cognitive. These results may be attributed, in part, to the specific age groups this study focused on and the nature of the three reading comprehension tests. The sample comprised of beginning, struggling readers who were in the process of mastering basic reading skills. Because the cognitive demands of the specific tests used were rather low, whereas the linguistic demands, and specifically the decoding demands were rather high, the children identified by these different tests were most likely struggling readers exhibiting decoding related difficulties.

Nevertheless, these results lead to some important conclusions. First, the results suggest that when processing speed is an integral component of a
test, as in the case of CBM-Maze, early naming speed and letter recognition difficulties can likely lead to fluency problems which, in turn, impact children’s ability to comprehend. Indeed, Katzir et al. (2006) have reported that a deficit in naming speed might be responsible for the slow recognition of letters and letter combinations in common orthographic patterns adversely affecting word recognition, with associated effects on dysfluent reading and comprehension. In addition, Bowers, Sunseth, and Golden (1999) have showed that children with naming speed deficits were less successful at finding letter strings in words than children with phonological deficits only, leading in turn to poor knowledge of word orthographic structure, dysfluent word reading, and thus poor comprehension. Simply put, for children who experience difficulties with letter identification and struggle with decoding words or read very slowly, the information in the text is not easily accessible (see also Perfetti, 2007).

Second, the results confirm that decoding skills make a substantial contribution toward explaining variation in reading comprehension performance in the early elementary school years. Phonological ability and rapid automatized naming, in particular, were shown to relate to the variation in performance obtained in reading comprehension more than other linguistic skills such as orthographic processing and letter identification, or even cognitive abilities such as information processing, attention or executive functioning. These results can be explained in part by a careful analysis of the component processes decoding depends upon. These are phonological decoding and orthographic processing, the two routes also related to the paths of the dual route model of reading (e.g., Coltheart, Rastle, Perry, Langdon, & Ziegler, 2001). Phonological decoding supports the reader to sound out words using grapheme-phoneme correspondences. Cumulative research in languages with scripts that are highly transparent such as Greek (Papadopoulos, Kendeou, & Spanoudis, 2012), Finnish (Torppa et al., 2014), or German (Wimmer, Landerl, & Frith, 1999) has concluded that phonological decoding constitutes the primary route to successful word reading. However, for reading to be fast, phonological decoding is not by itself adequate. Orthographic processing is also necessary for larger word units to be processed more automatically. However, orthographic processing depends upon RAN and children with slow naming speed process individual letters in a word too slowly to enable associations between the letters to be formed, thereby hindering the formation of good-quality representations of orthographic patterns that commonly occur in written text (e.g., Bowers & Wolf, 1993; Conrad & Levy, 2007; Georgiou, Papadopoulos, Zarouna, & Parrila, 2012). These poor-quality orthographic
reading comprehension tests and poor readers

representations, in turn, hinder development in vocabulary, semantic processing, and thus, reading comprehension (McCarthy, Hogan, & Catts, 2012). Therefore, with phonological ability and naming speed being not only precursors skills of reading comprehension but also significant sources of individual differences in reading that are used systematically for the taxonomy of poor readers (e.g., O’Brien, Wolf, & Lovett, 2012), it is not surprising that they emerged as the most significant deficits among our different groups of poor readers when compared to the control group.

Third, the results also suggest that word fluency problems seem to seriously disrupt reading flow and focus on comprehension, regardless of the test used to define poor comprehension. Indeed, all three groups of poor readers showed significant problems with word reading fluency in Grade 1 when compared to the control group. That fluency can potentially explain individual differences in reading comprehension is consistent with automaticity theory (LaBerge & Samuels, 1974) and the lexical quality hypothesis (e.g., Perfetti & Hart, 2001; Hart & Perfetti, 2008). Fluency frees important cognitive resources for the reader, resources that can be devoted to comprehension. Thus, reading fluency is the enabling link between decoding and reading comprehension.

At least two limitations of the present study are worth mentioning and addressing in future research. First, we didn’t include any predictors of language comprehension skills, such as syntactic or morphological awareness or inference skills that would have allowed us to test our poor readers’ profiles in a more comprehensive manner. Thus, future research should examine whether group differences in reading comprehension can be accounted for by differences in basic linguistic or cognitive processes other than those examined in the present study. Second, consistent with the “decoding” focus of this work, we selected and used reading comprehension tests that depended primarily on decoding skills. This approach also constrained the results we obtained, as we discussed above. Thus, future research could use tests that depend less on decoding and more on higher-order skills, such as inference generation and comprehension monitoring.

Concluding, the current set of findings contribute significantly to the literature by demonstrating similarities but also differences among children with reading comprehension difficulties (e.g., Nation, Clarke, Marshall, & Durand, 2004; Nation, Clarke, & Snowling, 2002; Yuill & Oakhill, 1991). The findings highlight that the use of different reading comprehension tests as diagnostic tools can result in different patterns of strengths and weaknesses in poor readers. Even though the findings highlighted...
weaknesses primarily in linguistic processes rather than cognitive processes, both sets of skills need to be considered in the early identification of reading difficulties.

Received April 18, 2013.
Revision accepted June 03, 2014.

REFERENCES


Cutting, L. E., & Scarborough, H. S. (2006). Prediction of reading comprehension: Relative contributions of word recognition, language proficiency, and other cognitive...
skills can depend on how comprehension is measured. *Scientific Studies of Reading*, 10, 277-299.


Reading comprehension tests and poor readers


