Learning Disabilities: A Contemporary Journal 6(1), 45-69, 2008

# Validation of Unsupervised Computer-Based Screening for Reading Disability in Greek Elementary Grades 3 and 4

**Athanassios Protopapas**[1]
*Institute for Language & Speech Processing*
**Christos Skaloumbakas**
*Athens Medical Pedagogical Center*
**Persefoni Bali**
*Institute for Language & Speech Processing*

*After reviewing past efforts related to computer-based reading disability (RD) assessment, we present a fully automated screening battery that evaluates critical skills relevant for RD diagnosis designed for unsupervised application in the Greek educational system. Psychometric validation in 301 children, 8-10 years old (grades 3 and 4; including 288 from the general school population and 13 from a clinical sample), indicated that computer-based screening can detect children likely to be diagnosed with RD (with 80-86% correct classification), using a linear discriminant function derived from measures taken without supervision within a 30-minute "computer game" interaction. We conclude that automated screening solutions constitute a feasible option in the context of a shortage of expert personnel.*

**Key Words: Reading Disability, Screening, Computer-Based Assessment, Primary Grades**

The development of written language skills is one of the most important functions of early schooling. Structured instruction aims at familiarizing children with the alphabetic principle and with the orthographic system of the language community and providing sufficient opportunity for practice so that children will move from a primarily oral mode of knowledge transfer to a written mode, better tailored to the demands of present-day educational and academic contexts. Consequently, individual variability in the facility with which written language skills are acquired is of principal concern for educators, as delays or deficits in attaining reading and spelling fluency may hold children back from achieving other academic goals and may even hinder their future personal and professional development (Blackorby & Wagner, 1996; Morrison & Cosden, 1997; Murray, Goldstein, Nourse, & Edgar, 2000).

Regardless of whether special causes are posited, or identified, to which failure of written language skill development can be attributed, it remains of great importance to identify and support children most in need of special intervention to improve their reading and spelling skills towards their age-appropriate levels. Wide-

1. Please send correspondence to: Athanassios Protopapas, ILSP/Athena, Artemidos 6 & Epidavrou, GR-15125 Maroussi, Greece; e-mail: protopap@ilsp.gr

range screening instruments are useful in this regard, insofar as reliable objective detection of children at risk for reading failure is possible. In educational systems with heightened awareness and special training of all educators to identify and refer children, special screening procedures may be redundant. However, when properly trained and experienced professionals are scarce and testing materials are unstandardized or unavailable, a standard screening procedure could greatly ease the task of detecting at-risk children. An automated, computer-based procedure might be especially useful, provided it meets the usual standards of psychometric adequacy and can be realistically applied in the school setting, given the level of technological competence of the available personnel (Cisero, Royer, Marchant, & Jackson, 1997; Protopapas & Skaloumbakas, 2007; Singleton & Vincent, 2004).

In the present article, we report the results of an effort to develop and validate a computer-based screening procedure for identifying children at risk for reading disability in the Greek educational system, grades 3 and 4. We first briefly document the selection of tasks most necessary for such assessment and review the current state of computer-based reading assessment. We then describe the educational context and previous work in Greece. Finally, we present and discuss the results of the first validation trial of our software.

## Reading Disability: Construct and Assessment

Despite much debate around the causes and surface indicators of reading disability, no consensus has yet been reached on a precise definition of the concept. The term *dyslexia*, generally used and still preferred within medical/biological research, appears more often in the British literature, whereas in North America the term *reading disabilities* prevails.

The concept of dyslexia, long associated with an aptitude-achievement discrepancy, has enjoyed wide acceptance by researchers and practitioners, as indicated by 94% of the U.S. states including the discrepancy in their criteria a decade ago (Mercer, Jordan, Allsopp, & Mercer, 1996). However, in the past 20 years, it has attracted severe criticism, eloquently summed up by Stanovich (1988, 2005). Criticism has been based, first of all, on strong evidence of similarity of cognitive, genetic, and neurologic characteristics between discrepant poor readers (i.e., those with a discrepancy between learning ability and academic progress) and nondiscrepant poor readers (now often called low achievers). Second, the notion of discrepancy was criticized for implying a valid qualitative difference between "dyslexics" and students who happen to be poor readers; a difference that remains unsubstantiated. Growing recognition that for more than 80% of children with learning disabilities, their primary problem is in reading (Lyon, 1995) has led to the adoption of the term *reading disability* as a theoretically neutral operational proxy of the learning problem.

Stanovich (1988) also proposed that a constellation of deficits may, in the case of dyslexia, center around a phonological "core," reinforcing a connection between reading disabilities and difficulties in phonological processing. Substantial cross-linguistic evidence has supported this connection both in the English language and in languages with more transparent orthographies as well, such as German (Landerl & Wimmer, 2000), Italian (Zoccolotti et al., 1999), and Spanish (Jimenez, 1997).

However, the nature and directionality of the relationship between reading disabilities and phonological processing remain unclear while its strength seems strongly modulated by age and degree of orthographic transparency.

Assessment of a reading disability (or dyslexia) is intertwined with developments in psychometrics, educational and psychological practices, and corresponding normative and legislative acts, as well as with the contemporary research agenda. Reading and spelling measures are essential surface indices of reading deficits. In addition, theoretical formulations such as the phonological core hypothesis stipulate that children with reading disabilities present phonological processing deficits. These deficits manifest themselves in the explicit manipulation of speech sounds ("phonological awareness"), in the short-term storage and recollection of verbal material, and in the rapid retrieval of phonological labels for highly familiar symbols (Goswami, 2003).

The most important and frequently used indices of reading disability involve reading isolated words and connected text. On the one hand, researchers agree that reading difficulties are present primarily at the word level, necessitating assessment of word reading accuracy and speed (typically in word lists). On the other hand, reading of connected text is important for establishing an ecologically valid measure of the academic deficit. Text reading puts the additional burden of syntactic and semantic processing on top of a presumably deficient word-reading process, thus potentially accentuating the word-level reading problem via cognitive resource straining. Therefore, it is not surprising that text reading may come out as a more significant indicator of word reading problems than single word reading in comparisons between children with reading disability and the general school population (Protopapas & Skaloumbakas, 2007).

Another commonly employed task in reading assessment is the reading of nonwords, the accuracy and speed of which measure the effectiveness and efficiency of the graphophonemic decoding route, relatively unaffected by lexical knowledge and sight reading (to the extent that the nonwords used do not strongly resemble existing words to be read by simple analogy). Additional measures typically include spelling of isolated words and connected text, lexical decision, and rapid automatized naming. Table 1 lists a representative set of measures employed by researchers working in European languages with transparent orthographies (for a discussion of orthographic transparency in European languages and its effects in early reading development, see Seymour, Aro, & Erskine, 2003).

## Computer-Based Reading Assessment

In recent years, a trend towards computer-based assessment of learning disabilities has become apparent internationally. Despite the popularity of the computer-based approach and several commercial products, the scientific literature on the subject remains limited, and only a few validation studies have been reported. Because of the scarcity of peer-reviewed publications on this topic, the following review of the software for identification of learning difficulties (reading disabilities in particular) is based in large part on Internet sources. We focus on computerized tests specifically designed for identification of reading disabilities, excluding tests converted from standard form to computer-based administration and also exclud-

Table 1

*Oral and Written Language and Cognitive Measures Used to Distinguish Between Dyslexic (Poor Readers) and Nondyslexic Children in Studies of Languages with Transparent Orthographies*

| | Word Reading Speed | Text Reading Speed | Pseudoword Reading Speed & Accuracy | Text Spelling | Word Spelling | Pseudoword Spelling | Pseudoword Repetition | Reading Comprehension | Listening Comprehension | Rapid Automatized Naming | Phonological Awareness | | Working (Short-Term) Memory |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | | | | | | | | | rime/onset | segm/blend | |
| Muller & Brady, 2001 (Finnish) | •• | | | | | | | | •• | •• | | •• | |
| Lehtola & Lehto, 2000 (Finnish) | •• | •• | •• | • | | •• | | | | | | | |
| Everatt, Smythe, Ocampo, & Gyarmathy, 2004 (Hungarian) | •• | | •• | | •• | •• | | | | • | •• | | •• |
| Wimmer, 1993 (German) | | •• | | | | •• | • | | | •• | • | •• | • |
| Wimmer, Mayringer, & Landerl, 2000 (German) | •• | •• | •• | | •• | | | | | •• | • | | |
| Spinelli et al., 2005 (Italian) | •• | •• | •• | •• | •• | •• | | | | | | •• | |
| Zoccolotti et al., 1999 (Italian) | •• | | •• | | | | | • (mixed) | | •• | | • | |

*Note.* • = Used; •• = Significant.

ing tests for the assessment of general abilities or for regular monitoring of students' (academic) progress.

Much work has been done in United Kingdom, where educational policies support extended use of computer-based assessment of learning difficulties at schools (Fawcett, Singleton, & Peer, 1998; Singleton & Vincent, 2004). Lucid Research (*About Lucid*, n.d.) has produced dyslexia assessment software for a wide range of ages (4 years old to adults) based on research conducted at the Department of Psychology of the University of Hull.

***Cognitive Profiling System***

The most widely used product is the Cognitive Profiling System (CoPS), a computerized psychometric assessment system for ages 4-8 aiming at "the early identification of cognitive strengths and limitations that affect the development of reading" (Singleton, Thomas, & Horne, 2000, p. 158). CoPS does not intend to replace traditional diagnosis but to identify children at risk of failure (underachievement) who need further assessment (Singleton et al., 2000; Singleton, Thomas, & Leedale, 2001). CoPS consists of eight tests in the form of games assessing sequential and associative memory, auditory and color discrimination, and phonological awareness (Singleton et al., 2000; Singleton et al., 2001).

The CoPS tests have been validated and standardized on a sample of 1,107 children and are currently used in schools in the U.K. and other English-speaking countries (*Lucid Cognitive Profiling System*, n.d.). Validation studies conducted on a sample of 421 children showed that CoPS scores at the age of 5 accounted for 31% and 50% of the variance in reading scores at the ages of 6 and 8 respectively. Discriminant-function analyses produced up to 100% correct classification of children's low reading ability (one standard deviation below the mean) versus good reading ability (at or above the mean), as measured by the Edinburgh Reading Quotient at age 8 years, based on CoPS measures taken at 5 years (Singleton et al., 2000). More recent documentation from Lucid (*Fact Sheet 4*, 2007) reports 96% overall correct prediction of poor reading skills, with 2.3% false positives and 17% false negatives, without specifying the exact measures, ages, or populations. Versions in other languages (Swedish, Norwegian, Italian, Arabic) are also in use or under development.

A brief assessment for the 4- to 15-year age range (Lucid Rapid Dyslexia Screening) tests phonological processing ability, auditory working memory, and visual memory skills (4-7 years) or phonic decoding skills (8-15 years) in about 15 minutes (Singleton, Horne, Leedale, & Thomas, 2003). Other products (LASS Junior and Secondary) are available for primary (8-11 years) and secondary (11-15 years) ages, the identification of specific learning difficulties, dyslexia, and screening of literacy problems, as well as for assessment of general ability and regular monitoring of progress in reading, spelling and memory skills (Horne, Singleton, & Thomas, 2001; Thomas, Horne, & Singleton, 2001). Each comprises eight modules (single-word and sentence reading, spelling, reasoning, auditory and visual memory, phonic skills and phonological processing) and lasts about 45 minutes. These tests have reportedly been validated and standardized on large representative U.K. populations (Horne et al., 2001; Thomas et al., 2001). Validation studies have showed "highly significant correlations between LASS Secondary measures and equivalent published conventional tests that are widely used by educational psychologists" (*Fact Sheet 4*, 2007). For people 16 years and older, computerized dyslexia screening tests measure word recognition and construction, working memory, and nonverbal reasoning, in about 20 minutes, emitting a probability of (having) dyslexia by reference to normative data (Singleton, Horne, Thomas, & Leedale, 2002). Other products by Lucid assess literacy, mathematics, communication skill, and personal and social development at school entry (Singleton, Horne, & Thomas, 1999).

***Cognitive Aptitude Assessment Software***

The Cognitive Aptitude Assessment Software (CAAS; *The Cognitive Aptitude Assessment Software*, n.d.), developed at the University of Massachusetts, is said to identify not only dyslexia but a wide range of learning disabilities, such as dyscalculia and attention deficit hyperactivity disorder (ADHD), by measuring the speed and accuracy of simple perception, letter identification, word, nonword and concept identification, sentence processing, number identification, and performance on addition, subtraction, multiplication and division problems.

Royer and Sinatra (1994) tested 112 children in grades 2 to 5 and found high reliability indices (.75-.80 for sentence verification and .88-.97 for response time measures) and good agreement with assigned reading-book level (as an index of validity). Subsequent validation studies with college students showed that the CAAS can distinguish between nondisabled, reading disabled, and learning disabled students; that it provides distinct profiles consistent with individual histories and standardized assessments; and that CAAS results can guide intervention (Cisero et al., 1997). A 20-minute screening version of CAAS is available for children in grades 3 through 8 (*Reading Success Lab*, n.d.).

***Comprehensive Diagnostic Assessment of Reading Difficulties***

The Comprehensive Diagnostic Assessment of Reading Difficulties (CDARD), based on the CAAS, aims to identify reading disorders at different levels (elementary, middle school, and high school/adult) by testing simple response, letter naming, word and nonword naming, sentence understanding, word meaning, listening, and reading comprehension (The Research-Based Comprehensive Diagnostic, n.d.). For both CAAS and CDARD, task items appear on the computer screen, and the student responds into a microphone so that reaction time can be automatically measured. However, an adult supervisor is required for providing instructions, supervising, and scoring the responses.

***Dyslexia Screener***

The Dyslexia Screener (*Dyslexia Screener*, n.d.) is a computer-based assessment designed to identify pupils (ages 4-19) with reading difficulties. Aiming at distinguishing between poor reading ability and dyslexia, it provides a profile of the student as well as suggestions for further assessment or intervention. It evaluates nonverbal reasoning, verbal comprehension, perceptual speed, phonological processing, reading (word recognition and comprehension), and spelling (letter recognition, word segmentation, and proofing) (Turner, 2004).

***InCAS***

In a similar vein, InCAS assessment software aims to identify reading problems and spelling difficulties, providing a profile of strengths and weaknesses for each student (ages 5-11). Specifically, InCAS assesses word recognition, reading, spelling, picture vocabulary, nonverbal ability, mental arithmetic, general mathematics, and attitudes, including self-concept (Merrell & Tymms, 2004).

***Lexia Reading Tests***

Lexia reading tests aim to highlight children' needs and help teachers identify specific areas where practice or instruction is needed. The Lexia Quick Reading Test (10 minutes) evaluates decoding skills, word attack strategies, and automaticity with sight words (*Lexia Quick Reading Test*, n.d.). The more extensive Lexia Comprehensive Reading Test (20-30 minutes) also assesses phonological awareness and reading comprehension and produces detailed reports for each skill to be used for progress monitoring (*Lexia Comprehensive Reading Test*, n.d.). Lexia software is only minimally computer-based, in that a teacher or other adult is needed to administer the tests, judging each response and manually keying in the corresponding data.

***Computerized Monitoring of Early Reading Skills***

Computerized Monitoring of Early Reading Skills (CMERS) is a computer-based adaptive test of reading skills for early readers in kindergarten through third grade (DeGraff, 2005). It measures phonemic awareness, phonics, reading fluency, vocabulary, and passage comprehension. An administrator gives instructions and technical support to the child, who navigates through the test and makes answer selections.

CMERS is not yet commercially available, but it has been thoroughly assessed for convergent and discriminant validity compared to conventional testing batteries. Data from about 120 children in each of grades K, 1, and 3 showed similar performance on CMERS and conventional measures in terms of classification performance. Specificity and positive predictive value in detecting children at risk (based on the Stanford Achievement Test or the Florida Comprehensive Assessment Test percentile cutoff criteria) exceeded 90% and 50%, respectively, for first grade, and 80% and 30%, respectively, for third grade. Sensitivity was 1.00 for first grade and .90 for third grade.

***Spelling Performance Evaluation for Language and Literacy***

Finally, focusing more specifically on spelling performance, Spelling Performance Evaluation for Language and Literacy (SPELL) is an assessment tool that may be used as part of a diagnostic battery, because it provides information on students' spelling error patterns and makes specific recommendations for intervention (Learning by Design, n.d.).

In summary, computerized assessment tools for identification of learning or reading disabilities differ in the skills measured, the age targeted, and the results they provide. None of them provides a clinical diagnosis. Rather, they may detect specific signs of reading disability and/or provide a profile of a student's strengths and weaknesses as compared to appropriate normative samples. Not all tests are adaptive or visually attractive (e.g., using graphical interfaces, animation, cartoon characters) to sustain children's interest and motivation. Further, none of the software programs presented above uses a thematic "scenario" linking the tasks together or presenting a concept or story to motivate children to participate. Psychometric validation information is not available for several tests; however, data from the few well-studied tests suggest that computerized assessment is a viable approach worth further study and development. Finally, large differences between tests in automatization and administration requirements were noted, ranging from minimally engaging the computer, only to display stimuli, to fully automatic administration, scoring and reporting procedures.

## Computer-Based Screening for Reading Disability in Greek

The Greek educational system is characterized by a significant lack of school-based services for children with specific educational needs. Despite recent legislation presumably catering precisely to such needs (Education of persons with special educational needs, 2000), in practice the structures of the Ministry of Education have been slow to develop, are severely understaffed and underfunded, and are typically staffed with educators or medical personnel who have little training or expertise in special education such as specific learning disabilities. Waiting periods of several months for assessment are not uncommon, and in the short time allotted for assessment of a child, the professionals authorized to carry out the assessment lack standardized materials or guidelines on which to base their clinical judgments. State-approved instruments for the diagnosis of specific learning disabilities include the Greek version of the Wechsler Intelligence Scales for Children-III (WISC-III) (Georgas, Paraskevopoulos, Bezevegis, & Giannitsas, 1997) and a testing battery called "AthenaTest" (Paraskevopoulos, Kalantzi-Azizi, & Giannitsas, 1999) for which no diagnostic criteria are provided and no reliability or validity data are available. As a result, diagnoses of dyslexia are typically given without objective measures of reading performance, based at best on uncontrolled and unvalidated custom-made materials by individual psychologists or special educators.

In this context, a validated computer-based screening tool might prove useful in at least detecting the children in the general population who might need extra attention in the form of professional assessment and individualized educational programs. Lacking standardized outcomes measures to compare against, Protopapas and Skaloumbakas (2007) developed "eMaDys," a computerized screening battery. eMaDys was tested against consensus expert clinical judgments formed on the basis of a custom set of reading assessment instruments examining reading accuracy, fluency, and comprehension, among other related skills such as spelling and phonological awareness. After establishing that experts relied on consistent and reasonable criteria from the clinical assessment to form their diagnostic judgment, a linear discriminant-function analysis determined an optimal application of the software measures for classification of children as at risk for reading disability or not. Classification rates (sensitivity as high as 90% with positive predictive value exceeding 50%; replicated with an independent validation sample) showed that the approach of automated computer-based screening is viable and potentially useful, especially against the backdrop of lack of services as described.

Because of the shortage of expert personnel, an important consideration in developing eMaDys was that it be entirely automated, requiring no intervention or supervision, and functioning independently of the potential availability of technologically minded or special education personnel. This requirement imposed significant restrictions on the kinds of tasks that can be tested; for example, precluding any tests with spoken or handwritten responses. Moreover, to guarantee that every child would be able to complete the tests without the need for assistance and without prior experience with computers, it was deemed necessary to exclude use of the keyboard, and to design all interaction using the mouse. Fortunately, use of the mouse has been found to be extremely simple and rapid to learn even for second-grade children with no prior computer experience.

An additional constraint, arising in part from the requirement for unsupervised administration, concerns sustaining the motivation and attention of children throughout the assessment. For this purpose, tests were given the form of (very simple) "computer games," in which a cartoon character guided the children through the tasks and received help from them in sorting out words, spellings, sounds, pictures, and so on. Observation of the children during testing indicated that the goal of keeping them focused was attained, so the results were considered reliable and useful (for the derivation of a classification function).

eMaDys was designed for children in seventh grade because this is a critical point in the Greek educational system. That is, at this point students move from the relatively lenient grading system of the elementary grades to a much more diverse and demanding secondary education environment, and they have to face written final examinations for the first time. Being legally entitled to oral (instead of written) testing if diagnosed with a learning disability, many children are referred to the state-approved diagnostic centers at this time to claim their benefits. Thus, even though it is rather late to diagnose and potentially treat a significant difficulty with written language (or other skill), the age of 12-14 is the one for which most referrals are seen for educational assessment.

In recent years, however, awareness of specific learning disabilities has increased in Greece, along with growing recognition of the significance of early intervention. Awareness stems in part from the aforementioned legislative mandate (Education of persons with special educational needs, 2000) calling for early identification and treatment of learning problems in the school population in conjunction with growing interest among researchers in the subject of reading disabilities.

In addition, most elementary schools in Greece have by now acquired at least some computers, rendering feasible a computer-based screening test for younger ages, where it might prove most useful. Based on the experience from the development of eMaDys and the results of its validation, an automated screening battery was thus developed for children in grades 3 and 4. Earlier grades were avoided in this effort for the following reasons. First, skill performance at grades 1 and 2 might be still rapidly developing, and hence unstable (Shaywitz, Escobar, Shaywitz, Fletcher, & Mahuck, 1992, found that only 28% of the children identified as reading disabled in grade 1 were still identified as such in grade 3). Second, executive functions required for carrying out the assessment tasks are highly variable and context-dependent (Zelazo, Carter, Resnick, & Frye, 1997), potentially diminishing the reliability of the screening results. Finally, younger children might have greater difficulty at a first contact with a computer, jeopardizing the validity of the data.

## Purpose of the Study

Below we report the findings from a validation study of the newly developed software screening battery for reading disability, "VLEMA," designed for children in Greek grades 3 and 4. Due to the lack of commonly accepted diagnostic guidelines and objective standardized outcomes measures, we first examine the differences between the general school population and a clinical population of children professionally diagnosed with specific reading disability (dyslexia). Our aim was to objectively define a subgroup of schoolchildren from the general population as at risk for dyslexia on the basis of internationally accepted types of assessment measures and then to examine the classification performance of the software in detecting these children.

In setting up a new screening instrument for earlier grades, we have capitalized on our experience from eMaDys. As mentioned, an indispensable constraint for automated assessment concerns complete automatization, thus precluding any forms of user input other than mouse clicks. Spoken or written input is cumbersome to enter and impossible to score automatically, whereas keyboard-based input requires extensive experience and facility with computers in order not to constitute an impediment to performance. This greatly restricts the types of assessment tasks that can be included in an automated screening battery.

In addition, the entire screening process must be completed in a short time to avoid student fatigue and to be easily integrated with competing educational activities in the school setting. Therefore, the number and the duration of assessment tasks are also severely constrained. Thus, only the most successful tasks from eMaDys were selected for the new instrument on the basis of their contribution to the linear discriminant function that correctly detected seventh-grade RD children (see Protopapas & Skaloumbakas, 2007). These tasks, described under Materials below, were retained because of their theoretical and practical importance for RD detection and because similar clinically administered measures discriminated between RD children and the general school population in grades 3 and 4, as shown in the Results section.

## Method

### *Participants*

Data from 288 children (137 boys and 151 girls) from the general school population, constituting the "school sample," are reported below (after excluding 20 incomplete datasets). This sample included 145 children from grade 3 and 143 from grade 4. The students were recruited in 10 elementary schools from the Athens metropolitan area, spanning a wide range of estimated socio-economic conditions, and one school from Thessaloniki. We only tested children whose parents returned a signed consent form. In agreement with teachers' comments, we estimate this self-selection procedure to have resulted in a higher proportion of below-average performing children than would be expected in a more representative sample from the general population.[1]

In addition, 13 children (7 boys and 6 girls) made up the "clinical sample." These children were attending grades 3 and 4 and were recruited, independently from the school sample, at the Athens Medical Pedagogical Center, where they sought assessment for learning disabilities due to low academic performance. All the children in the clinical sample were diagnosed with specific reading disability (dyslexia) follow-

[1]Apparently, more parents of well-performing children were reluctant to consent to missing instruction time due to prolonged teacher strikes earlier in the academic year.

ing comprehensive clinical evaluation on the basis of poor reading and spelling performance despite average or higher intelligence.

The mean age of the children was 103.0 months (*SD* = 4.2) for grade 3 and 114.3 months (*SD* = 4.1) for grade 4.

## *Materials*

### *Clinical Assessment*

The traditional clinical assessment included the following tasks (for more information on the measures, see Protopapas & Skaloumbakas, 2007).

***Pseudoword repetition.*** Number of incorrectly pronounced items in a list of 20 pseudowords pronounced by the experimenter one by one for the child to repeat.

***Pseudoword reading.*** Total reading-aloud time and number of incorrectly read items for a list of 20 pseudowords 2-5 syllables long.

***Word reading.*** Total reading-aloud time and number of incorrectly read items for a list of 84 words 1-7 syllables long.

***Passage reading and comprehension.*** Total reading-aloud time and number of incorrectly read words in three age-appropriate passages 40-76 words long. Each passage was followed by 2-3 open-ended comprehension questions, scored according to pre-determined criteria.

***Spelling to dictation.*** Number of incorrectly spelled items in a list of 21 isolated words and number of spelling errors in a dictated 33-word passage.

***Phoneme deletion.*** Number of incorrect responses from a set of 22 pseudowords pronounced by the experimenter to be first repeated and then produced with a specified phoneme omitted.

***Speech discrimination.*** Number of incorrect "same/different" responses in the 32-pseudoword-pair speech discrimination subscale of the "Athena" test for learning disabilities (Paraskevopoulos et al., 1999).

***Nonverbal intelligence.*** Number of correct items in lists A-D of the Standard Progressive Matrices (Raven, 1976).

***Digit span.*** Total number of correctly reproduced sequences (raw score) in the digit span subscale from the Greek standardized version of WISC-III (Georgas et al., 1997).

***Arithmetic.*** Number of correct responses (plus five; raw score) in the arithmetic subscale from the Greek standardized version of WISC-III (Georgas et al., 1997).

***Rapid color and letter naming (RAN).*** Time to name 60 color patches (red, green, blue, yellow, and brown in equal proportions) and time to name 60 letters (α, δ, κ, λ, σ) presented in three columns of 20 items each.

### *Computer-Based Assessment*

Children were administered the computer-based assessment battery VLEMA, developed in Asymmetrix Toolbook, a multimedia authoring program for the PC. VLEMA was designed according to the same principles as eMaDys (see Protopapas & Skaloumbakas, 2007), namely, simplicity and automation, leading to ease of use. Briefly, *simplicity* refers to the presence of only functional elements in the tasks, no distractions, strictly linear navigation, and no options or choices for the user other than for responding to the assessment items. All user input is based on the computer mouse; a virtual keyboard displayed on the screen allows selection of letters using the mouse when necessary. *Automation* means that no human intervention is needed at any stage. Installation proceeds automatically without the need for any choices, and test administration and scoring require no supervision. All that is needed from the supervisor (teacher) is to ensure that the child has the headphones on correctly (to hear the recorded oral instructions included with the software) and sits comfortably in front of the computer.

VLEMA includes the following four tests.

***Reading speed and comprehension.*** Each of 10 passages (22-53 words long; $M = 39.8$) were presented on screen for a length of time determined by the child (clicking with the mouse to begin and end text display) and were subsequently followed by a choice of four line drawings, one of which depicted the situation described in the passage. Reading time and comprehension score (number of correct image selections) were recorded.

***Letter span.*** Random sequences of Greek uppercase single consonants were presented at a rate of about one per second for the child to reproduce by clicking with the mouse on the correct letters in the correct order. Sequence length progression and stopping rule were modeled after the standard digit span task (starting length was two items). The total number of correctly reproduced sequences was recorded.

***Pseudoword spelling to dictation.*** Twenty pseudowords, 1-3 syllables long, were presented auditorily, one at a time, for the child to spell by clicking on the appropriate letters on the screen. Any legal spelling retaining the phonological identity of the pseudoword was considered correct. The number of correct items was recorded.

***Word identification.*** Thirty sets of four alternative spellings and a line drawing depicting a word meaning were presented for the child to select the correct spelling of the depicted word. The number of correct responses and the mean response time were noted.

### *Testing Procedures*

The clinical assessment was administered individually in a quiet room at school (for the school population) or at the children's hospital (for the clinical population) by a specially trained psychologist or graduate student. Breaks were given to allow the children to rest when tired. Testing time did not exceed two 40-minute class periods. All tasks in which children responded verbally were tape recorded, and scoring was later verified from the tapes.

Group testing with the software was determined by the availability of computers at the school lab with the restriction that children not be able to see one another's screen. Closed-type headphones were always used, through which spoken instructions were delivered by the software at the beginning of each task. Two easy examples, with decreasing guidance, preceded each task. Testing time did not exceed one class period and was, for most children, substantially shorter.

The order of testing batteries was not systematically manipulated or counterbalanced, determined primarily by scheduling convenience as indicated by the school authorities.

## Results

Table 2 shows the main performance characteristics for both grades (school population only) in the clinical assessment and computer-based tasks. Internal con-

Table 2

*Possible Range, Reliability, Critical Percentiles, and Difference Between General and Dyslexic Groups for Clinical Assessment Measures*

| | Possible Range | | Reliability | Grade 3 Percentiles | | | Grade 4 Percentiles | | | General vs. Dyslexic | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | Poorest | Best | (Cronbach's α) | 10 | 50 | 90 | 10 | 50 | 90 | *F* | *p* | $\eta^2$ | Cohen's *d* |
| Pseudoword repetition errors | 20 | 0 | .72 | 9 | 4 | 1 | 8 | 3 | 1 | 17.25 | .000 | .055 | 1.40 |
| Pseudoword reading errors | 20 | 0 | .83 | 12 | 8 | 2 | 12 | 6 | 0 | 14.78 | .000 | .047 | 1.22 |
| Pseudoword reading time (s) | | | | 110 | 62 | 43 | 93 | 54 | 40 | 2.26 | .134 | .007 | .36 |
| Word reading errors | 84 | 0 | .91 | 23 | 12 | 3 | 20 | 7 | 1 | 8.35 | .004 | .027 | 1.02 |
| Word reading time (s) | | | | 243 | 151 | 108 | 201 | 123 | 83 | 8.48 | .004 | .028 | .79 |
| Text reading errors | | | | 11 | 3 | 0 | 9 | 2 | 0 | 8.90 | .003 | .029 | 1.00 |
| Text reading time (s) | | | | 225 | 132 | 87 | 161 | 97 | 71 | 16.20 | .000 | .051 | 1.23 |
| Reading comprehension | 0 | 20 | .44 | 6 | 11 | 15 | 8 | 11 | 14 | 2.52 | .113 | .008 | .45 |
| Text spelling errors | | | | 16 | 6 | 2 | 12 | 2 | 0 | 18.70 | .000 | .059 | 1.36 |
| Word spelling errors | 21 | 0 | .87 | 14 | 7 | 2 | 11 | 3 | 0 | 18.76 | .000 | .059 | 1.34 |
| Phoneme deletion errors | 22 | 0 | .86 | 16 | 11 | 4 | 16 | 9 | 3 | 12.55 | .000 | .040 | 1.06 |
| Speech discrimination errors | 32 | 0 | .78 | 12 | 7 | 2 | 10 | 4 | 1 | 14.40 | .000 | .046 | 1.27 |
| Raven's SPM raw score | 0 | 48 | .89 | 13 | 22 | 34 | 15 | 24 | 36 | .04 | .851 | .000 | .05 |
| WISC-III digit span raw score | 0 | 30 | | 8 | 11 | 14 | 8 | 12 | 15 | 9.39 | .002 | .030 | .96 |
| WISC-III arithmetic raw score | 0 | 30 | | 9 | 14 | 17 | 11 | 15 | 19 | .41 | .520 | .001 | .22 |
| RAN colors time (s) | | | | 77 | 56 | 45 | 74 | 54 | 41 | .87 | .353 | .003 | .32 |
| RAN letters time (s) | | | | 82 | 56 | 40 | 67 | 48 | 38 | 6.36 | .012 | .021 | .88 |
| Reading comprehension | 0 | 9 | .68 | 2 | 6 | 8 | 3 | 7 | 9 | 1.32 | .252 | .004 | .27 |
| Passage reading time (s) | | | .93 | 46 | 26 | 14 | 38 | 23 | 14 | 3.43 | .065 | .011 | .42 |
| Pseudoword spelling accuracy | 0 | 20 | .75 | 2 | 8 | 14 | 3 | 10 | 16 | 9.05 | .003 | .029 | .92 |
| Word identification accuracy | 0 | 30 | .86 | 14 | 23 | 28 | 18 | 27 | 30 | 6.22 | .013 | .020 | .78 |
| Word identification time | | | .86 | 10 | 7 | 4 | 9 | 6 | 4 | 21.58 | .000 | .067 | 1.36 |
| Letter span | | | | 2 | 4 | 6 | 2 | 4 | 6 | 1.17 | .280 | .004 | .34 |

*Note.* Clinical assessment measures (top section); computer-based measures (bottom section). Grade 3, *N* = 145; grade 4, *N* = 143; general population group, *N* = 285; dyslexic group, *N* = 16.

sistency (Cronbach's α) was computed for accuracy measures with a fixed number of items. Due to severe deviations from normality, measures were transformed as appropriate (typically, times via inverse and errors via square root). Normalized (*z* score) values were then computed separately for each grade to allow merging of data from the two grades. All the following analyses were conducted on the transformed and normalized values.

The general school population is expected to include children at all levels of performance, even children who would receive a diagnosis of reading disability if professionally evaluated, Therefore, it is not a homogeneous group of "average" performaners. However, by comparing the school sample to the clinical sample, we can expect to identify the relative importance of each measure for the identification of reading disability, because the clinical sample is composed entirely of children with reading disability whereas the school sample represents the general population.

Table 2 also shows the results of ANOVA comparing these two groups on each measure.[2] It is clear that the measures of the clinical assessment best distinguishing the two groups ($d > 1.0$) were pseudoword repetition, reading accuracy and speed (especially for passages), spelling accuracy, and the phonological measures of phoneme deletion and speech discrimination. Digit span narrowly missed the arbitrary effect-size cutoff ($d = .96$), followed by RAN letters ($d = .88$). The computer-based measures most useful for this group distinction were word-picture identification accuracy and (especially) speed, along with pseudoword spelling.

Table 3 lists the correlation coefficients between the measures from the clinical and the computer-based assessment (for the school sample only). The amount of variance in each clinical measure that can be accounted for by the computer-based measures (in linear regression with all computer-based measures entered together as independent predictors) is also shown. That the computer-based measures are well chosen to identify children with reading difficulties is evidenced by the fact that the amount of accounted-for variance roughly follows the importance of each clinical measure for the school-clinical distinction. Note that more than half of the variance of several important clinical measures, including spelling and reading measures, can be accounted for by the computer-based measures.

The critical question is whether this set of computer-based measures can identify the specific individuals in the general school population who are most likely to be diagnosed with reading disability. To examine this issue, it is necessary to form a clear, independent criterion of nonoverlapping group membership. Because our school sample includes children with reading problems, conducting a discriminant analysis of the computer-based measures of the school vs. clinical grouping is not likely to produce clear results. Instead, we can first form two distinct groups by using the clinical assessment measures, entirely independently of the computer-based assessment. These clinical assessments constitute an objective tool on the basis of which an expert is expected to provide a diagnosis of reading disability. Therefore, we can assume that clear-cut cases of reading disability should be separable from the general population by a linear combination of these measures. This assumption is empirically supported by previous findings (for Greek seventh grade) showing that

[2]Three children from the school group were included with the clinical group instead for this analysis, because they were certifiably diagnosed with dyslexia and performed very poorly on our reading tests.

Table 3
*Absolute Values of Pearson Product-Moment Correlation Coefficients Between Clinical Assessment Measures and Computer-Based Measures*

| | RC | RT | PS | WI | WIT | LS | $R^2$ |
|---|---|---|---|---|---|---|---|
| Pseudoword repetition errors | .33 | .12 | .29 | .33 | .20 | .29 | .19 |
| Pseudoword reading errors | .31 | .24 | .30 | .45 | .22 | .31 | .26 |
| Pseudoword reading time | .18 | .39 | .21 | .26 | .37 | .22 | .25 |
| Word reading errors | .45 | .30 | .37 | .59 | .33 | .38 | .47 |
| Word reading time | .34 | .51 | .27 | .51 | .41 | .32 | .49 |
| Text reading errors | .35 | .21 | .30 | .49 | .31 | .29 | .31 |
| Text reading time | .36 | .57 | .32 | .62 | .44 | .36 | .63 |
| Reading comprehension | .35 | .15 | .28 | .30 | .12 | .21 | .18 |
| Text spelling errors | .39 | .35 | .42 | .64 | .37 | .38 | .53 |
| Word spelling errors | .40 | .30 | .40 | .65 | .39 | .34 | .52 |
| Phoneme deletion errors | .37 | .21 | .34 | .50 | .18 | .38 | .32 |
| Speech discrimination errors | .35 | .01 | .36 | .34 | .09 | .33 | .23 |
| Raven's SPM raw score | .34 | .02 | .31 | .36 | .05 | .35 | .22 |
| WISC-III digit span raw score | .25 | .21 | .33 | .40 | .18 | .38 | .25 |
| WISC-III arithmetic raw score | .23 | .14 | .27 | .39 | .24 | .28 | .20 |
| RAN colors time | .18 | .11 | .21 | .26 | .15 | .20 | .10 |
| RAN letters time | .22 | .27 | .23 | .45 | .27 | .32 | .27 |

*Note*. Coefficients of .25 or higher are significant at $p$ <.00005. The right-hand column contains the cumulative proportion of variance in the corresponding clinical assessment measure that can be accounted for by all computer-based measures entered together in linear regression analysis. RC: Reading comprehension; RT: Reading time; PS: Pseudoword spelling; WI: Word identification accuracy; WT: Word identification speed; LS: Letter span. $N$=288.

children unambiguously classified (as ND or RD) are identified 100% by the same set of clinical assessment measures (Protopapas & Skaloumbakas, 2007).

A linear discriminant-function analysis was conducted on the sample groups using as predictors the nine clinical assessment measures that showed a large effect size in separating the two groups ($d > 1.0$ in Table 2). The full school sample ($N$ = 288) was contrasted with the full clinical sample ($N$ = 13). Table 4 shows the results of the analysis, and Table 5 lists the classification outcomes using equal prior probabilities (.50) for the two groups. The function constant was .016. Function values at group centroids were -.074 for the school sample and 1.633 for the clinical sample. Examination of the resulting discriminant-score distribution of the school and clinical samples indicated that the clinical sample overlapped with 25% of the school sample, in agreement with the impression that our self-selected school sample included a relatively high proportion of children with poor academic abilities.

Distinct subgroups of the school sample were then formed on the basis of cutoff scores on this linear discriminant function (20th percentile and poorer vs. 30th percentile and higher) resulting in a new grouping with 57 children in the RD range (henceforth "RD") and 202 children outside of it (nonimpaired; henceforth "NI"). This derived grouping could be discriminated at 99% (98% in leave-one-out cross-validation) in a confirmatory linear discriminant-function analysis using the same nine predictors (all children were classified correctly except for 3 NI children misclassified as RD). The derived RD subgroup of the school sample did not differ significantly from the clinical sample in any measure except reading comprehension, $F(1, 67) = 10.10$, $p = .003$, in which the derived RD group performed worse than the clinical sample.

A linear discriminant-function analysis on the NI group ($N = 202$) vs. the clinical sample ($N = 13$) using all computer-based measures as predictors, with equal prior probabilities (.50) for the two groups, resulted in 86.5% (84.7% cross-validated) correct classification, based primarily on word identification response time, word identification accuracy, and pseudoword spelling. Table 6 (top) shows the linear discriminant function information, and Table 7 (top) shows the results of the classification. The function constant was .307. Function values at group centroids were -.122 for the NI group and 1.892 for the clinical group.

Finally, the ability of computer-based measures (all entered together) to discriminate the derived RD vs. NI grouping was tested in an additional LDFA analysis. Discrimination performance, again with equal prior probabilities, was 80.7% (79.2% cross-validated). The linear discriminant-function information and classification results for this analysis are shown in Tables 6 (bottom) and 7 (bottom). The function constant was .109. Function values at group centroids were -.368 for the NI group and 1.305 for the clinical group. Based on these classification results for the last analysis, we can compute the clinically important indexes of sensitivity (75%), specificity (82%), and positive predictive value (54%).

Table 4

*Results of Linear Discriminant-Function Analysis Contrasting the School Sample with the Clinical Sample on the Basis of the Nine Measures of the Clinical Assessment with Effect Size Greater Than or Equal to 1.0 in a Comparison of the Two Groups by ANOVA*

| | CDFC | SCDFC | PWGC |
|---|---|---|---|
| Pseudoword repetition errors | .374 | .366 | .690 |
| Pseudoword reading errors | .562 | .547 | .639 |
| Word reading errors | .672 | .654 | .480 |
| Text reading errors | .025 | .025 | .496 |
| Text reading time | .282 | .275 | .669 |
| Text spelling errors | .234 | .228 | .718 |
| Word spelling errors | .318 | .307 | .720 |
| Phoneme deletion errors | .062 | .061 | .589 |
| Speech discrimination errors | .271 | .265 | .630 |

*Note.* School sample: $N = 288$; clinical sample: $N = 13$. CDFC: (unstandardized) canonical discriminant-function coefficients; SCDFC: standardized canonical discriminant-function coefficients; PWGC: pooled within-groups correlations between discriminating variables and standardized canonical discriminant function. Absolute values of the actual coefficients are shown.

Table 5

*Classification Outcomes of Linear Discriminant-Function Analyses Contrasting the School Sample with the Clinical Sample on the Basis of the Nine Measures of the Clinical Assessment with Effect Size Greater Than or Equal to 1.0 in the Comparison of the Two Groups by ANOVA*

| | Predicted Group | |
|---|---|---|
| Actual Group | School | Clinical |
| School | 230 | 58 |
| Clinical | 1 | 12 |

*Note.* School sample: *N* = 288; clinical sample: *N* = 13. The actual grouping (rows) is contrasted with the grouping predicted from the linear discriminant function (columns).

Table 6

*Results of Linear Discriminant-Function Analyses Using the Derived Subgroups of the School Sample*

| | CDFC | SCDFC | PWGC |
|---|---|---|---|
| Reading comprehension | .089 | .083 | .759 |
| Passage reading time | .014 | .012 | .582 |
| Pseudoword spelling accuracy | .432 | .416 | .540 |
| Word identification accuracy | .518 | .451 | .372 |
| Word identification time | .759 | .697 | .270 |
| Letter span | .009 | .009 | .247 |
| Reading comprehension | .131 | .122 | .522 |
| Passage reading time | .117 | .107 | .291 |
| Pseudoword spelling accuracy | .351 | .333 | .590 |
| Word identification accuracy | .754 | .654 | .873 |
| Word identification time | .221 | .203 | .349 |
| Letter span | .147 | .145 | .460 |

*Note.* Top: Contrasting the derived nonimpaired (NI) subgroup of the school sample (*N* = 202) with the clinical sample (*N* = 13) on the basis of the computer-based assessment measures. Bottom: Contrasting the derived nonimpaired (NI) subgroup (*N* = 202) with the reading disabled (RD) subgroup (*N* = 57) of the school sample on the basis of the computer-based assessment measures. CDFC: (unstandardized) canonical discriminant-function coefficients; SCDFC: standardized canonical discriminant-function coefficients; PWGC: pooled within-groups correlations between discriminating variables and standardized canonical discriminant function. Absolute values of the actual coefficients are shown.

Table 7
*Classification Outcomes of Linear Discriminant-Function Analyses Using the Derived Subgroups of the School Sample*

| | Predicted Group | |
|---|---|---|
| Original Group | NI | RD |
| NI | 175 | 27 |
| Clinical | 2 | 11 |
| NI | 166 | 36 |
| RD | 14 | 43 |

*Note*. The original grouping (rows) is contrasted with the grouping predicted from the linear discriminant function (columns). Top: Contrasting the derived nonimpaired (NI) subgroup of the school sample ($N$ = 202) with the clinical sample ($N$ = 13) on the basis of the computer-based assessment measures. Bottom: Contrasting the derived nonimpaired (NI) subgroup ($N$ = 202) with the reading-disabled (RD) subgroup ($N$ = 57) of the school sample on the basis of the computer-based assessment measures.

## Discussion

We have presented a study of computer-based screening for reading disability in third and fourth grade of Greek schools, in which four measures taken without supervision in the form of "computer games" in about half an hour are compared to a much larger set of measures taken in structured clinical assessment interviews individually administered in more than double the time. The internal reliability of the computer-based measures is comparable to that of the clinically administered measures. Measured against a clinical diagnostic criterion (for the clinical sample) or a statistical grouping based on the clinical assessment (for the school sample), it was found that the great majority of RD children could be successfully identified on the basis of the computer-based measures. The percentage of children identified as possibly RD by the computer-based procedure who are likely to be RD according to the clinical assessment is not less than 50%, similar to that obtained previously for seventh-grade children (Protopapas & Skaloumbakas, 2007).

Thus, the classification performance of VLEMA is similar to previous results reported for Greek (grade 7; Protopapas & Skaloumbakas, 2007) and English students (grades K to 3; DeGraff, 2005). The proportion of variance in traditional measures accounted for by the computer-based measures was higher than that reported for CoPS (cumulative $R^2$ not exceeding .45; Singleton et al., 2000), but it should be noted that CoPS scores were used to predict reading scores taken one year later and not at the same time as in our study. The substantially higher classification performance reported for CoPS may be largely due to the very wide separation between their poor and good readers. That is, Singleton et al. (2000) removed 34% of their sample by excluding all children with Edinburgh Reading Quotients up to one standard deviation below the mean, whereas we imposed only minimal separation by excluding 10% of the children (between the 20th and 30th percentile on the linear discriminant function from the clinical assessment). Our effect sizes for the comparison between children with reading disability and the general population ranged between .27 and 1.36, similar to those reported for CAAS (for college students; Cisero et al., 1997; Table 9, p. 617).

The clinical measures on which RD children were found to be furthest from the general population mean were, not surprisingly, reading fluency (speed), reading and spelling accuracy, and the phonological measures (pseudoword repetition, phoneme deletion, and speech discrimination). The finding of the importance of text reading speed is in agreement with previous studies in Greek (with younger children; Porpodas, 1999) and other languages with consistent orthographies (e.g., Lehtola & Lehto, 2000; Spinelli et al., 2005; Wimmer, Mayringer, & Landerl, 2000, see Table 1), in which poor reading fluency was found to be among the most important indices of reading difficulties. Similarly, the importance of spelling accuracy is consistent with corresponding findings in the German language, which is comparable to Greek in being complex to spell despite high forward consistency (that is, graphophonemic rules permit correct reading of most words, whereas correct spelling requires specific word knowledge). The importance of the phonological measures in these ages, in agreement with findings in a host of other languages (e.g., Caravolas & Volín, 2001; Jiménez González & Hernández Valle, 2000; Mayringer & Wimmer, 2000), stands in contrast to their relative insignificance for older Greek children (Protopapas & Skaloumbakas, 2007) and remains consistent with a view of a gradually diminishing role of phonological awareness in the further development of reading expertise (McBride-Chang, 1995; Shaywitz, 2003).

Turning to the individual computer-based measures, it was interesting, and somewhat unexpected, to find that the reaction time to word identification was by far the strongest discriminant variable in the detection of RD children. Even though pseudoword spelling to dictation, a largely phonological measure, ranked second in effect size (and canonical DF coefficient), further attesting to the importance of phonological processing for RD identification, the importance of word identification speed points to the development of sight vocabulary as a sensitive and critical area of reading skill that remains deficient in reading disability. It is notable that word identification speed does not correlate more highly than word identification accuracy with the clinical assessment measures, which is to be expected on the basis of the hypothesized similarity between the measures. Therefore, the observed importance of word identification speed for RD identification does not reflect some failure of word identification accuracy to capture relevant aspects of written word processing as measured by standard reading and spelling tasks. Rather, it is possible (though in need of further substantiation) that this measure taps more directly the quality, or efficiency, of lexical representations that is relevant for successful reading, including the strength of the connections among their semantic, orthographic, and phonological components (as posited by the "lexical quality hypothesis;" Perfetti & Hart, 2001, 2002).

It was largely expected that reading comprehension measures would not distinguish between NI and RD, and this was confirmed in the analyses. This finding is in agreement with the corresponding results from the 7th-grade study (Protopapas & Skaloumbakas, 2007) and with a host of recent reports from other languages on the dissociability of word-level reading skill from text-level comprehension, both in the general population (Cain, Oakhill, & Bryant, 2004; Oakhill, Cain, & Bryant, 2003) and in special groups (Cain, Oakhill, & Bryant, 2000; Hoover & Gough, 1990; Megherbi & Ehrlich, 2005). The lack of a significant difference between NI and RD was observed for both the clinical and the computer-based measures of comprehension; therefore, it does not reflect the failure of either measure to assess the construct of interest. The correlation between computer-based and clinical assessment of comprehension is moderate and comparable in size to other significant correlations (see Table 3) between computer-based reading comprehension score and clinical measures of reading and spelling, further attesting to the construct validity of the computer-based reading comprehension measure, and increasing our confidence in the conclusion that this skill is not relevant for the identification of word-level reading disability (dyslexia).

In contrast, it was unexpected to find that computer-based passage reading time would prove useless for the NI-RD distinction. This finding does not reflect a failure of the automated procedure to assess the intended construct, because (a) the correlation coefficients between computer-based reading speed and clinically administered tests of word and text reading speed are among the highest (in Table 3); and (b) the internal consistency of the computer-based reading speed measure is the highest of all (including the clinical measures; see Table 2). It seems that there is something about reading speed that is well captured by the computer-based measure, which is not important for RD detection, and something about reading speed that is missed by the computer-based measure, which is very important. There are two obvious differences between the clinical and the computer based measures of reading speed: (a) in the clinical assessment measures, children read aloud whereas in computer-based assessment they read silently; and (b) computer-based reading is self-paced, initiated and ended by the child's mouse click. These differences mean that, in the computer-based task (but not in the clinical assessment), children may skip words they find too difficult or they may move on without delaying to pronounce easy yet irrelevant words. Moreover, in the computer-based task (and not in the clinical assessment), children may take their time to reread parts of the passages (or entire passages) they find more demanding, in order to increase their comprehension (picture selection) performance.

Therefore, the computer-based measure of reading speed apparently does not measure text reading fluency, but text processing efficiency, which includes much more than simply going through the word sequence and converting the orthographic representations to phonological ones. Thus, one plausible interpretation of the findings is that the word-level skill of going through the word sequence, necessarily included in the process, causes the correlation with clinical measures of reading speed, whereas the higher-level executive skill of freely processing the text for comprehension dominates the measure and dissociates it from the NI-RD distinction. This interpretation is strengthened by the observation that word identification speed, a purer measure of word-level processing efficiency, does strongly predict the NI-RD distinction.

The main question in the present study was whether automatized computer-based assessment procedures can reliably identify children who are most in need of special attention to be referred for comprehensive evaluation and educational support for their academic weaknesses. The results suggest that this is a feasible aim and that useful application of computer-based screening is possible within the expected limits arising from the more restricted set of measures and less controlled procedure in which they are derived. Our results are comparable to previously reported findings for computer-based reading disability assessment, and the reliability of screening using VLEMA is satisfactory as long as one is prepared to accept the low positive predictive value of the procedure as a trade-off for the low cost, brief duration, and great ease of use.

The interpretation of the present results is limited by the lack of a comprehensive evaluation and an independent expert diagnostic judgment for each of the participating children, which means that our origin-based or statistically derived groupings can only be approximate, nonhomogeneous, and likely including children meeting potential exclusion criteria due to poor intelligence, attention deficits, and so on. Therefore, the performance indices obtained for the computer-based screening are best considered low estimates of the true value of the automated procedure, which could only be obtained in the (impractical) case of large, clean groups of individually examined children. Additional important limitations of this study include the small number of children tested, which may have distorted our estimates of group performance due to inadequate sampling of the target populations, and the lack of properly controlling demographic factors of potential significance such as gender and socioeconomic status.

The small number of measures employed in the computer-based assessment is also limiting our ability to identify tasks with optimal features as far as practical identification of children with reading problems is concerned. Finally, our nonhomogeneous small groups and potentially biased sampling may also have adversely affected the distributions of our measures and estimates, thus reducing the robustness of the statistical procedures employed and the reliability of the results. Therefore, the current study is best seen as constituting a feasibility study rather than a definitive report on the outcomes and recommended practices regarding computer-based screening for middle primary-school grades.

In conclusion, we find computer-based screening for reading disability to be a feasible, psychometrically adequate, practical alternative to traditional clinical assessment for purposes of initial screening and referral. Its highest potential use would be in cases where available expert personnel are insufficient to address the needs of the general school population. Standardization on a large-scale sample representative of the general population is necessary before widespread adoption, however. In addition, both providers and recipients of screening services, including parents, educators, and special education administrators, must be fully informed and well aware of the fact that computer-based assessment alone cannot provide comprehensive evaluation, let alone diagnosis, and that automated screening solutions can only constitute a first step aiming at selecting an at-risk subset of children and producing referrals for much-needed proper professional services.

## References

*About Lucid.* (n.d.). Retrieved February 9, 2007, from http://www.lucid-research.com/about.htm.

Blackorby, J., & Wagner, M. (1996). Longitudinal postschool outcomes of youth with disabilities: findings from the National Longitudinal Transition Study. *Exceptional Children*, *62*, 399-414.

Cain, K., Oakhill, J., & Bryant, P. (2000). Phonological skills and comprehension failure: A test of the phonological processing deficit hypothesis. *Reading and Writing*, *13*, 31-56.

Cain, K., Oakhill, J., & Bryant, P. (2004). Children's reading comprehension ability: Concurrent prediction by working memory, verbal ability, and component skills. *Journal of Educational Psychology*, *96*, 31-42.

Caravolas, M., & Volín, J. (2001). Phonological spelling errors among dyslexic children learning a transparent orthography: The case of Czech. *Dyslexia*, *7*, 229-245.

Cisero, C. A., Royer, J. M., Marchant, H. G., & Jackson, S. J. (1997). Can the computer-based academic assessment system (CAAS) be used to diagnose reading disability in college students? *Journal of Educational Psychology*, *89*, 599-620.

*The cognitive aptitude assessment software.* (n.d.). Retrieved January 26, 2007, from http://www.cognitive-aptitude-assessment-software.com

DeGraff, A. J. (2005). *Monitoring growth in early reading skills: Validation of a computer adaptive test.* Unpublished doctoral dissertation, Department of Psychology, University of Florida.

*Dyslexia screener.* (n.d.). Retrieved January 19, 2007, from http://www.nfer-nelson.co.uk/education/resources/dyslexia_screener/dyslexia_screener.asp?css=1

Education of persons with special educational needs. (2000, 14 April). *Efimerida tis kiverniseos* [*Official Government Gazette*], *78*(tefxos A). (Law 2817).

Everatt, J., Smythe, I., Ocampo, D., & Gyarmathy, E. (2004). Issues in the assessment of literacy-related difficulties across language backgrounds: A cross-linguistic comparison. *Journal of Research in Reading*, *27*, 141-151.

*Fact sheet 4.* (2007, January). Retrieved April 29, 2007, from www.lucid-research.com/documents/factsheets/FS04_ResearchAndEvaluationStudies.pdf.

Fawcett, A. J., Singleton, C. H., & Peer, L. (1998). Advances in early years screening for dyslexia in the United Kingdom. *Annals of Dyslexia*, *48*, 57-88.

Georgas, D. D., Paraskevopoulos, I. N., Bezevegis, I. G., & Giannitsas, N. D. (1997). *Elliniko WISC-III: Wechsler klimakes noimosinis gia paidia* [Greek WISC-III: Wechsler Intelligence Scales for Children]. Athens: Ellinika Grammata.

Goswami, U. (2003). Why theories about developmental dyslexia require developmental designs. *Trends in Cognitive Sciences*, *7*, 534-540.

Hoover, W. A., & Gough, P. B. (1990). The simple view of reading. *Reading and Writing: An Interdisciplinary Journal*, *2*, 127-160.

Horne, J., Singleton, C. H., & Thomas, K. V. (2001). *LASS secondary teacher's manual* (2nd ed.). Beverley, East Yorkshire, UK: Lucid Research Creative Ltd.

Jimenez, J. E. (1997). A reading-level design study of phonemic awareness underlying reading disabilities in a transparent orthography. *Reading and Writing: An Interdisciplinary Journal*, *9*, 23-40.

Jiménez González, J. E., & Hernández Valle, I. (2000). Word identification and reading disorders in the Spanish language. *Journal of Learning Disabilities*, *33*, 44-60.

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.

*Learning by design - products - SPELL.* (n.d.). Retrieved February 13, 2007, from http://www.learningbydesign.com/products/spell.htm.

Lehtola, R., & Lehto, J. E. (2000). Assessing dyslexia in Finnish high school students: A pilot study. *European Journal of Special Needs Education*, *15*, 255-263.

*Lexia comprehensive reading test.* (n.d.). Retrieved February 12, 2007, from http://www.lexialearning.com/forschools/products/compreadingtest.php

*Lexia quick reading test.* (n.d.). Retrieved February 12, 2007, from http://www.lexialearning.com/forschools/products/quickreadingtest.php

*Lucid Cognitive Profiling System.* (n.d.). Retrieved February 27, 2007, from http://www.lucid-research.com/sales/esales.htm?category_id=31&product_id=181.

Lyon, G. R. (1995). Research initiatives in learning disabilities: contributions from scientists supported by the national institute of child health and human development. *Journal of Child Neurology*, *10*, S120-S126.

Mayringer, J., & Wimmer, H. (2000). Pseudoname learning by German-speaking children with dyslexia: Evidence for a phonological learning deficit. *Journal of Experimental Child Psychology*, *75*, 116-133.

McBride-Chang, C. (1995). What is phonological awareness? *Journal of Educational Psychology*, *87*, 179-192.

Megherbi, H., & Ehrlich, M.-F. (2005). Language impairment in less skilled comprehenders: The on-line processing of anaphoric pronouns in a listening situation. *Reading and Writing*, *18*, 715-753.

Mercer, C. D., Jordan, L. A., Allsopp, D., & Mercer, A. R. (1996). Learning disabilities definitions and criteria used by state education departments. *Learning Disabilities Quarterly*, *19*, 217-232.

Merrell, C., & Tymms, P. (2004). *Diagnosing and remediating literacy problems using INCAS software: Identifying reading and spelling difficulties and providing help* (Tech. Rep.). Durham. Retrieved April 3, 2007, from http://www.pipsproject.org/Documents/CEM/publications/downloads/CEMWeb039%20Incas.pdf

Morrison, G. M., & Cosden, M. A. (1997). Risk, resilience, and adjustment of individuals with learning disabilities. *Learning Disability Quarterly*, *20*, 43-60.

Muller, K., & Brady, S. (2001). Correlates of reading performance in a transparent orthography. *Reading and Writing: An Interdisciplinary Journal*, *14*, 757-799.

Murray, C., Goldstein, D. E., Nourse, S., & Edgar, E. (2000). The postsecondary school attendance and completion rates of high school graduates with learning disabilities. *Learning Disabilities Research & Practice*, *15*, 119-127.

Oakhill, J. V., Cain, K., & Bryant, P. E. (2003). The dissociation of word reading and text comprehension: Evidence from component skills. *Language and Cognitive Processes*, *18*, 443-468.

Paraskevopoulos, I. N., Kalantzi-Azizi, A., & Giannitsas, N. D. (1999). *AthinaTest diagnosis diskolion mathisis* [AthenaTest for the diagnosis of learning difficulties]. Athens: Ellinika Grammata.

Perfetti, C. A., & Hart, L. (2002). The lexical quality hypothesis. In L. Verhoeven, C. Elbro, & P. Reitsma (Eds.), *Precursors of functional literacy* (pp. 189-213). Amsterdam/Philadelphia: John Benjamins.

Porpodas, C. D. (1999). Patterns of phonological and memory processing in beginning readers and spellers of Greek. *Journal of Learning Disabilities*, *32*, 406-416.

Protopapas, A., & Skaloumbakas, C. (2007). Computer-based and traditional screening and diagnosis of reading disabilities in Greek. *Journal of Learning Disabilities*, *40*, 15-36.

Raven, J. (1976). *Standard progressive matrices.* San Antonio, TX: The Psychological Corporation.

*Reading success lab.* (n.d.). Retrieved January 25, 2007, from www.readingsuccesslab.com.
*The research-based comprehensive diagnostic assessment for reading difficulties module.* (n.d.). Retrieved February 12, 2007, from

http://www.readingsuccesslab.com/FullAssessment.html.
Royer, J. M., & Sinatra, G. M. (1994). A cognitive theoretical approach to reading diagnostics. *Educational Psychology Review*, *6*, 81-113.
Seymour, P.H.K., Aro, M., & Erskine, J. M. (2003). Foundation literacy acquisition in European orthographies. *British Journal of Psychology*, *94*, 143-174.
Shaywitz, S. E. (2003). *Overcoming dyslexia: A new and complete science-based program for reading problems at any level.* New York: Alfred Knopf.
Shaywitz, S. E., Escobar, M. D., Shaywitz, B., Fletcher, J. M., & Mahuck, R. (1992). Evidence that dyslexia may represent the lower tail of a normal distribution of reading ability. *New England Journal of Medicine*, *326*, 145-150.
Singleton, C. H., Horne, J., Leedale, R. C., & Thomas, K. V. (2003). *Lucid rapid dyslexia screening: Administration, interpretation and teaching guide* (1st ed.). Beverley, East Yorkshire, UK: Lucid Research Ltd.
Singleton, C. H., Horne, J., & Thomas, K. (1999). Computerised baseline assessment of literacy. *Journal of Research in Reading*, *22*, 67-80.
Singleton, C. H., Horne, J., Thomas, K. V., & Leedale, R. C. (2002). *LADS version 1.0 administrator's manual.* Beverley, East Yorkshire, UK: Lucid Innovations Ltd.
Singleton, C. H., Thomas, K., & Horne, J. (2000). Computer-based cognitive assessment and the development of reading. *Journal of Research in Reading*, *23*, 158-180.
Singleton, C. H., Thomas, K. V., & Leedale, R. C. (2001). *CoPS 1 cognitive profiling system manual* (2nd ed.). Beverley, East Yorkshire, UK: Lucid Research Ltd.
Singleton, C. H., & Vincent, D. (2004). Assessing literacy: current challenges and issues. *Journal of Research in Reading*, *27*, 113-117.
Spinelli, D., De Luca, M., Di Filipppo, G., Mancini, M., Martelli, M., & Zocolotti, P. (2005). Length effect in word naming in reading: role of reading experience and reading deficit in Italian readers. *Developmental Neuropsychology*, *27*, 217-235.
Stanovich, K. E. (1988). Explaining the difference between the dyslexic and the garden-variety poor reader: The phonological-core variable-difference model. *The Journal of Learning Disabilities*, *21*, 590-604.
Stanovich, K. E. (2005). The future of a mistake: Will discrepancy measurement continue to make the learning disabilities field a pseudoscience? *Learning Disability Quarterly*, *28*, 103-106.
Thomas, K. V., Horne, J., & Singleton, C. H. (2001). *LASS junior teacher's manual* (1st ed.). Beverley, East Yorkshire, UK: Lucid Research Ltd.
Turner, M. (2004). *The dyslexia screener – A resource for schools and colleges.* Retrieved January 19, 2007, from http://www.nfer-nelson.co.uk/articles/dyslexiascreener.asp.
Wimmer, H. (1993). Characteristics of developmental dyslexia in a regular writing system. *Applied Psycholinguistics*, *14*, 1-33.
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.
Zelazo, P. D., Carter, A., Resnick, J. S., & Frye, D. (1997). Early development of executive function: A problem solving framework. *Review of General Psychology*, *1*, 198-226.
Zoccolotti, P., De Luca, M., Di Pace, E., Judica, A., Orlandi, M., & Spinelli, D. (1999). Markers of developmental surface dyslexia in a language (Italian) with high grapheme-phoneme correspondence. *Applied Psycholinguistics*, *20*, 191-216.

***Author's Note***

The testing software was developed by programmer G. Koulafetis using artwork by graphic designer G. Magakis. Development was supported in part by grant 02PRAXE39/2003 from the Greek Secretariat for Research and Technology to A.P.

We thank C. Tsagaraki, S. Gerakaki, S. Alexandri, A. Grigoriadou, Y. Vogindroukas, E. Selini, S. Haskou, S. Drakopoulou, K. Koutsi, L. Hadjiantoniou, and D. Kasselimis for administering the tests.

Preliminary results from the work reported here were presented at the 2006 Learning Disabilities Worldwide conference in Burlington, Massachusetts.

Received, March 3, 2007
Revised, July 20, 2007
Accepted, October 1, 2007