Journal of Educational Psychology
1997, Vol. 89, No. 4, 599–620

0022-0663/97/$3.00

Can the Computer-Based Academic Assessment System (CAAS) Be Used to Diagnose Reading Disability in College Students?

Cheryl A. Cisero
Wayne State University

James M. Royer
University of Massachusetts

Horace G. Marchant III and Stanley J. Jackson
Westfield State College

Four studies are presented that evaluate the validity of the computer-based academic assessment system (CAAS) as a diagnostic technique for identifying specific reading disability in college students. CAAS assesses component reading skills using computer-presented reading tasks that measure speed and accuracy of performance. CAAS validity was evaluated against 4 requirements of a reading diagnostic: The technique must (a) be valid for the purpose of identifying reading disability, (b) provide data that are consistent with theories of reading disability, (c) provide information about the specific nature of the student's problem, and (d) provide prescriptive information that leads to the development of interventions that alleviate students' learning problems. Study results show that the CAAS system can satisfy all 4 criteria.

Researchers and practitioners in the field of developmental dyslexia, or specific reading disability, have begun to change their view on how reading disability should be defined and diagnosed. Traditionally, reading disability, and learning disabilities in general, have been defined using exclusionary definitions, which characterize the learning disability by enumerating various factors that do not contribute to the difficulty (e.g., low intelligence, sensory impairments, lack of educational opportunity, or socioeconomic/cultural background) rather than specifying the difficulties that are responsible for the learning problem. These exclusionary definitions also served as the basis for learning disability diagnosis. Specifically, the exclusion of low intelligence as a factor contributing to learning disabilities was transformed into an IQ-achievement discrepancy technique, which has become the most widely used method of diagnosing learning disabilities. Simply put, the IQ-achievement discrepancy identifies students as learning disabled if they have "normal" intelligence, as measured by a standardized IQ test, but perform significantly below what is expected from their IQ in one or more academic achievement domains.

More recently, researchers and practitioners in the field of specific reading disability have begun to recognize the inadequacies of both the exclusionary definitions and the IQ-achievement discrepancy method. The vagueness of exclusionary definitions regarding the nature of the reading difficulty makes accurate diagnosis nearly impossible and offers little help to educators in developing effective instructional interventions (Lyon, 1995). Moreover, the validity of the IQ-achievement discrepancy as a diagnostic technique has been called into question on conceptual, empirical, and statistical grounds (e.g., Evans, 1990; Fletcher et al., 1994; Morrison & Siegel, 1991; Reynolds, 1981, 1985; Shepard, 1980; Siegel, 1989, 1992; Stanovich, 1991a, 1991b; Stanovich & Siegel, 1994).

This growing consensus among professionals regarding the shortcomings of both the definition of reading disability and the primary diagnostic technique has served as a catalyst for reform in the areas of definition and diagnosis. Progress has already been made in developing a new definition of specific reading disability that operationalizes the disability using inclusionary, rather than exclusionary, criteria. In a recent paper, Lyon (1995) suggested that an inclusionary definition should be one that specifies the characteristic deficits of reading-disabled individuals as proposed by theory and supported by research evidence. The working definition of reading disability put forth by the Orton Dyslexia Society Research Committee in 1994 serves as an example of a research-based inclusionary definition. This definition characterizes reading disability as a specific language-based disorder of constitutional origin that is primarily due to an impairment in phonological processing ability that affects the development of decoding ability and word recognition skill (Lyon, 1995).

Cheryl A. Cisero, College of Education, Wayne State University; James M. Royer, Department of Psychology, University of Massachusetts; Horace G. Marchant III and Stanley J. Jackson, Department of Psychology, Westfield State College.

Parts of Experiments 1, 2, and 3 reported in this article were presented at the annual meeting of the American Educational Research Association. Experiment 2 was reported in a dissertation by Cheryl A. Cisero, which was submitted in partial fulfillment of the requirement for a doctoral degree at the University of Massachusetts, Amherst, February 1996. The dissertation was completed under the direction of James M. Royer.

We thank Joe Shinn, Sue Krieg, and Dan Neuner for their invaluable contribution to this research.

Correspondence concerning this article should be addressed to Cheryl A. Cisero Durwin, 345 College of Education, Wayne State University, Detroit, Michigan 48202. Electronic mail may be sent via Internet to cdurwin@cms.cc.wayne.edu.

New diagnostic techniques, if they are to be an improvement over conventional methods, must be developed along the same lines as a new definition of specific reading disability. That is, a diagnostic technique for identifying specific reading disability should be based on a theory regarding the skills involved in reading acquisition and the difficulties that impede reading development and should be supported by research evidence. One assessment technique that fits this description is the Computer-Based Academic Assessment System (CAAS).

The CAAS system, which was developed by Royer and colleagues (Royer & Sinatra, 1994; Sinatra & Royer, 1993) at the University of Massachusetts, measures the speed and accuracy of performance on computer-presented tasks that were designed to assess component processes involved in reading (explained in more detail in the next section). In this article, we present four studies to evaluate the validity of the CAAS system as an instrument for diagnosing specific reading disability in college students. The validity of the CAAS system is evaluated against four criteria of a valid diagnostic system proposed by Royer and Sinatra (1994). These criteria are discussed in detail following a description of the CAAS system.

The Computer-Based Academic Assessment System

The CAAS is a battery of computer-presented tasks in which stimuli are presented on a computer screen and examinees make responses by either pressing a button or speaking into a microphone. The tasks were designed, based on theory, to reflect cognitive component processes believed to be involved in the development of skilled reading. Therefore, an explanation of the theory that drives the CAAS system is essential for understanding the purpose of CAAS as a reading diagnostic.

The CAAS system was developed based on a cognitive-developmental theory of reading that has evolved over a number of years through the contributions of many researchers and scholars (e.g., Anderson, 1983; Fodor, 1983; Forster, 1979; Perfetti, 1988, 1992; Stanovich, 1986, 1990; van Dijk & Kintsch, 1983; see Royer & Sinatra, 1994, for a more detailed discussion of cognitive-developmental theory). The theory suggests that skilled reading is based on the development of a number of component processes. Development of successful reading begins with two enabling (or prereading) skills: phonological awareness and the ability to identify letters (Adams, 1990; Bradley & Bryant, 1985; Stanovich, 1986). The acquisition of phonological awareness (the ability to recognize that speech can be divided into constituent sounds) and letter identification skill enables the child to discover the alphabetic principle that printed letters are represented by speech sounds. This discovery serves as the basis for the acquisition of decoding ("sounding out") skill and subsequent development of word recognition skill (Liberman & Liberman, 1990). Word recognition in skilled readers is characterized as fast, impervious to higher level cognitive processes (e.g., strategy use and prior knowledge), and nearly load-free (i.e., using little working memory capacity). Another characteristic of skilled reading is the automatic activation of a word's meaning once the word has been recognized. The automatic identification of words and activation of their meanings allow the reader to input and hold enough words in working memory (which has limited capacity and a short duration) so that a meaningful chunk can be processed before the words begin to decay from memory (Royer & Sinatra, 1994). This is important for the next component, semantic processing, whereby the reader interprets text information by connecting it to related information in memory.

Development of skilled reading, therefore, depends on the acquisition of the following processes: phonological awareness, letter identification, word identification, concept (word meaning) activation, and semantic processing. Failure to develop a component process could result in a bottleneck that impedes the development of all subsequent higher level processes. Skilled reading not only depends on the development of the above component processes, but also depends on adequate prior knowledge and strategic processing. The reason is that the interpretation of text information during semantic processing requires relating printed information to prior knowledge as well as using metacognitive and reading strategies to assist in text interpretation, particularly when the reader experiences difficulty in understanding the text. Consequently, according to cognitive-developmental theory, there are three potential sources of reading failure: a failure to develop any one of the component skills, insufficient prior knowledge, and inadequate metacognitive and strategic processing (Royer & Sinatra, 1994).

The CAAS system was developed to assess whether children had adequately developed the component processes described above (except for phonological awareness, which cannot as yet be assessed using the CAAS system given the auditory nature of the task). The CAAS reading battery was originally created for elementary school children. In recent years, an adult version of the CAAS reading battery has also been developed (as well as elementary reading and math tasks in Spanish, a middle school reading battery and elementary-level and middle-school–level mathematics batteries). The elementary- and adult-level CAAS reading tasks are relevant for this article, and therefore are described next. Examples of items in each of the tasks can be found in the Appendix.[1]

For each task, the CAAS program randomly (without replacement) presents an examinee with a subset of items from the larger pool of items in the task. The CAAS system automatically records the examinee's response time and accuracy in the case of button-press tasks. In microphone tasks, the computer records response time when the examinee speaks into the microphone and the examiner records response accuracy on-line by pressing a *correct* or *incorrect* button on a scorer's box connected to the computer.

Simple Response Time Task

The simple response time task is a measure of the speed and accuracy of responding to nonverbal stimuli. The task

[1] A complete list of stimuli for the tasks in the CAAS battery used in the experiments in this article is available on request.

contains eight "***" items and eight "+++" items, and the CAAS program selects 15 stimuli for presentation. When the microphone is used, examinees respond by saying "star" or "plus." When button presses are used, examinees respond to the displays of stars with one button and pluses with another.

The task is the first one presented in the battery and serves to acclimate the examinee to the testing situation. The simple response time task also provides a baseline indication of the speed with which examinees can perform a simple perceptual task. As will be seen later, this baseline indication of perceptual processing is useful in differentiating between various types of learning disability.

Letter Identification Task

In this task, the examinee says the name of an uppercase or lowercase letter that appears on the computer screen. The task contains the 52 upper- and lowercase letters of the alphabet, of which the CAAS program selects 18 letters for presentation.

An early version of the CAAS system used a variation of the Posner letter-match task (e.g., Posner, Boies, Eichelman, & Taylor, 1969) that required examinees to decide whether pairs of letters had the same or different names. Three different types of stimuli were presented: letter pairs that were physically identical (AA), same-name letter pairs (Aa), and physically different letter pairs (AB). Physically identical (AA) and same-name (Aa) pairs have the same name, and physically different (AB, bh) pairs have different names. This task contains 40 identical-match items, 40 name-match items, 40 uppercase nonmatch items, and 40 lowercase nonmatch items. Ten of each of the item types are presented, and the examinee presses one button for same-name items and another button for different-name items.

Elementary-Level Word and Pseudoword Naming Tasks

The word naming task is a measure of word recognition skill. Examinees pronounce single three- to six-letter words into a microphone. All words have been reported to be familiar to at least 80% of Grade 4 students (Dale & O'Rourke, 1976). The pseudoword task serves as a measure of decoding ability. Stimuli are pronounceable nonwords that have been derived from the real words by changing one letter in each real word. The word and nonword tasks contain 60 three-, four-, five-, and six-letter words (for a total of 240). Ten words or nonwords at each difficulty level (for a total of 40 items) are presented to an examinee.

Adult-Level Word and Pseudoword Naming Tasks

The adult word-naming task consists of words at six difficulty levels: one-syllable words with regular and irregular spelling patterns, two-syllable regular and irregular words, and three-syllable regular and irregular words. Half of the words at each difficulty level are low frequency (defined as less than 50 occurrences per million) and half high frequency (over 100 occurrences per million; Francis & Kucera, 1982). Pseudowords were constructed from the real words by changing one letter per syllable.

The word tasks contain 40 words or nonwords from each of the six difficulty levels (a total of 240 words). The CAAS program randomly selects 4 words from each of the first two easy categories (one-syllable regular and irregular words) and 8 words from the remaining four categories (regular and irregular two-syllable and three-syllable words), for a total of 40 words presented to an examinee.

Elementary-Level and Adult-Level Category Match Tasks

The category match task measures the ability to activate concepts in semantic memory. Examinees are informed of the categories to be included in the task and are then presented with pairs of words. Examinees indicate whether or not the words belong to the same category. When button presses are used, examinees press one button if the words are from the same category and another button if the words are from different categories. When the microphone is used, examinees respond with "same" or "different."

For the elementary-level task, the categories are transportation, animals, fruits, body parts, and clothes, and the adult task uses politics, economy, and general science. A new version of the adult-level category task has recently been developed that presents synonym and antonym pairs because it appeared that the original task was serving more as a measure of prior knowledge than of concept activation.

The elementary-level category task consists of 60 pairs of words, of which half are from the same categories and half are from different categories. Twenty items are selected for presentation. The adult category task contains 72 word pairs, half from the same category and half from different categories, with 25 of the items selected for presentation to the examinee.

Elementary-Level and Adult-Level Semantics Tasks

The semantics task assesses the application of semantic knowledge in sentence processing using a variation of the cloze procedure. Examinees are presented with sentences that contain a blank and a word above and below the blank. Examinees indicate which of the two words (which vary in semantic appropriateness) best fits the sentence. When button presses are used, the examinee presses one button to choose the word above the blank and another button to choose the word below the blank. When the microphone is used, the examinee says the word into the microphone. Elementary-level and adult-level tasks are identical except that in the adult-level task the sentences are longer and the word choices are more complex vocabulary words. The elementary-level semantics task consists of 70 cloze sentences, 15 of which are presented to an examinee, and the adult semantics task contains 40 sentences, of which 18 are presented.

Phonological Processing Tasks

Three phonological processing tasks were added to the adult CAAS battery. These tasks are visually presented "phonological awareness" tasks that measure an examinee's ability to detect rhyme, initial phonemes, and final phonemes in pairs of words. In each task there are four item types: (a) words that share the target sound and spelling pattern, (b) words that share the target sound but not spelling pattern, (c) words that do not share the target sound but share spelling pattern, and (d) words that neither share target sound nor spelling pattern. For instance, in the rhyme task, two words that rhyme and are spelled similarly are *shoot/boot,* words that rhyme and are spelled differently are *shoot/fruit,* words that do not rhyme but have similar spelling patterns are *shoot/foot,* and words that neither rhyme nor share spelling pattern are *shoot/walk.* It should be noted that although these tasks are not a pure measure of phonological processing (as they are presented visually rather than aurally), the tasks do require phonological processing because examinees are asked to make judgments about whether the words have similar sounds.

The tasks contain one-syllable words that have been reported to be familiar to at least 80% of Grade 12 students (Dale & O'Rourke, 1976). Each of the phonological processing tasks contains 120 word pairs (30 from each of the four-item type categories). Examinees are presented with 10 pairs from each item type category (for a total of 40 items).

Criteria of a Successful Reading Diagnostic

In a recent article, Royer and Sinatra (1994) outlined four criteria for a successful reading diagnostic. They are: (a) the technique must be reliable and valid, (b) the patterns of performance obtained from the technique must be consistent with cognitive-developmental theory of reading, (c) the technique should provide information about the specific nature of the reading problem, and (d) diagnoses obtained from the technique must lead to prescriptive procedures that alleviate the reading problem. The authors then discussed research on the elementary-level CAAS system related to Criteria 1 and 2. For a more detailed discussion, see Royer and Sinatra (1994). The research is summarized briefly here.

The initial evaluation of the elementary-level CAAS system involved individually administering the elementary-level tasks to students from two elementary schools in Grades 3 through 5 and retesting students 1 year later. First consider Criterion 1, that the technique must be reliable and valid. Tasks were found to be highly reliable (reliability indices of the response time measures ranging from .88 to .97). With respect to validity of CAAS as a measure of reading skill, performance on CAAS tasks was related to three indices of reading skill: grade level, students' current reading book level, and teacher ratings of reading skill (Royer & Sinatra, 1994).

With respect to the second criterion, Sinatra and Royer (1993) found that patterns of performance on the CAAS system were consistent with cognitive-developmental theory. For instance, word recognition skills of older students were more strongly related to sentence processing than that of younger students. According to cognitive-developmental theory, beginning readers' word recognition processes are slow and involve the use of strategies and contextual information. More experienced readers have word identification processes that are becoming fast and automatic. Therefore, word identification performance at the level where students are beginning to develop automated skills makes a stable and consistent contribution to sentence comprehension (Royer & Sinatra, 1994).

The initial evaluation of the CAAS system suggested that the technique satisfied Criteria 1 and 2 of a successful reading diagnostic when used to assess reading competence in elementary school children. The research reported in the current article examines the extent to which the CAAS system can fulfill all four of Royer and Sinatra's (1994) criteria when used as a reading diagnostic with college students. Experiments 1 and 2 address the first criterion by evaluating a basic requirement of a valid reading diagnostic, that the technique differentiate reading disabled students from nondisabled students and from students with other disabilities. The first two experiments also provide evidence that is relevant to the issue of whether assessment results are consistent with theory (Criterion 2), though the nature of this question is slightly different than it was in Sinatra and Royer (1993). Sinatra and Royer asked whether CAAS results were consistent with cognitive-developmental theories of reading. In the experiments to be reported, the question is whether patterns of performance for disabled and nondisabled readers are consistent with expectations derived from theories of reading disability.

Experiments 1, 2, and 3 are also relevant to Royer and Sinatra's (1994) third criterion—whether the technique can provide information about the specific nature of the reading problem. However, this criterion is also slightly modified in the present research. Royer and Sinatra (1994) had suggested that a reading diagnostic must identify the specific nature of the problem being experienced. In the present research, we maintain this criterion but add the requirement that the nature of the problem identified by CAAS assessments should be different for college students with specific reading disabilities and college students with learning disabilities that do not exclusively affect reading performance. Finally, in Experiment 4 in this article we address the issue relevant to Criterion 4—whether the information provided by CAAS assessments can be used to devise interventions that alleviate the reading problems of reading-disabled students.

Before proceeding to the studies, we wish to point out that the research in this article not only represents a growing body of evidence on the CAAS system as a college-level reading diagnostic, but also represents the constraints we have faced in conducting learning disability research. For instance, in Experiments 1 and 3, during the earlier part of our research efforts, we were not allowed access to learning disabled students' records to verify their disabilities and to systematically classify them into groups with different types of disabilities. Therefore, the groups in these studies basically represent a school-identified sample. In Experiment 2, we were allowed access to student records but were rather naive in expecting that the records would contain all pertinent information regarding a student's disability, such as test scores and the history of a student's academic difficulties, that would be necessary in classifying and describing our sample. Needless to say, this was not always the case, as will become evident later.

The lack of data in student records not only represents an obstacle for our research but is also characteristic of the state of learning disability diagnosis at postsecondary institutions in general. Many learning disability centers at colleges and universities, such as the one where we conducted our research, do not have the necessary staff to handle assessment and diagnosis of learning disabilities and therefore require students to obtain evaluations from outside professionals. Consequently, the quality and the type of evaluation that students bring as documentation of a learning disability vary, and make it virtually impossible to classify students for research purposes. It should be emphasized, though, that this difficulty in learning disability identification at the postsecondary level makes the need for improved diagnostic techniques even stronger.

Experiment 1

A minimal requirement of a valid assessment technique for identifying specific reading disability is that it distinguish the performance of reading-disabled students from that of nondisabled students and students with other types of learning disabilities. Experiment 1 was a preliminary attempt at determining whether the CAAS system could make this distinction in college students using the elementary-level CAAS reading battery (the adult battery had not been developed yet).

Method

Participants. Twenty-eight students were recruited from Disabled Student Services (DSS) at a college in western Massachusetts. DSS is a small learning disability center, staffed mainly by a director and a counselor. Students are required to bring evidence of a disability to DSS, and then, based on the submitted documentation, staff makes a decision regarding the student's eligibility for learning disability services and the types of accommodations that will be provided.

After all data collection was completed in the present study, the director of DSS classified 19 of the 28 students as having a specific reading disability (RD), and 9 as having a generalized learning disability (LD) based on his clinical judgment of each student's disability. The director's clinical judgment was derived from interviews, test scores, other documentation, and personal experience with the students. Because the authors were not allowed access to student records, placement into RD and LD groups was based solely on the director's clinical judgment, and no other information about the participants can be provided. It is also worthwhile to mention that the DSS staff played no part in the studies reported in this article other than classifying students into diagnostic categories. Forty nondisabled students who reported having no reading difficulties were recruited from an introductory psychology course at the same college to serve as a comparison group.[2]

Tasks and procedure. All participants were individually administered the elementary-level CAAS reading battery (described earlier). The elementary-level battery consisted of the simple response-time task, letter match, and the elementary versions of the word-naming, pseudoword-naming, category match, and semantics tasks. This was an earlier version of the CAAS system that used button-press responses for the simple, letter, category, and semantics tasks. For each task, participants were told to respond as quickly and as accurately as possible. Participants were also given practice trials and were allowed to ask questions prior to each task. The total time for administration was approximately 30 min.

Results

Reliabilities for the CAAS tasks were not computed in the experiments in this article because the item sampling procedure resulted in a different test for each examinee. However, it should be noted that the tasks administered in Experiment 1 were essentially the same as those reported in Royer and Sinatra (1994) where the CAAS task reliabilities all exceeded .80.

Accuracy and response time scores of the nondisabled, RD, and LD groups are shown in Table 1. Response time data were examined in a multivariate analysis of variance with diagnostic group (nondisabled, RD, LD) as a between-subjects factor and task as a within-subject factor. Given that accuracy performance of the groups was near ceiling, nonparametric tests were used on the accuracy data as these tests are considered to be more powerful than the F test with samples having skewed distributions (Myers & Well, 1991). Kruskal-Wallis tests on each of the CAAS tasks revealed a significant difference among the groups only on the pseudoword task, $\chi^2(2, N = 68) = 6.20, p < .05$. The RD and LD groups were less accurate (87% and 89%, respectively) on pseudowords than the nondisabled group (92%).

With respect to the response time analysis, a significant main effect of task was found, $F(5, 325) = 549.46, MSE = .06, p < .001$. Table 1 shows that response time increased as a function of the complexity of the tasks. A significant effect of diagnostic group was also obtained in the response time data, $F(2, 65) = 5.67, MSE = .22, p < .01$. Overall, the nondisabled group was faster than the RD and LD groups.

A significant group by task interaction was also obtained for the response time data, $F(10, 325) = 3.43, MSE = .06, p < .001$. The nature of this interaction was that the RD and LD groups exhibited different patterns of response time performance on the tasks relative to the nondisabled group. Figure 1 displays the response time interaction in terms of percentile scores of the RD and LD groups that were converted from effect sizes. Effect sizes are calculated by subtracting the mean response time of each of the two experimental groups (RD and LD groups) from the mean response time of the control group (nondisabled group) and dividing by the standard deviation of the control group. The

[2] No attempt could be made in either Experiment 1 or 2 to match disabled and nondisabled students on ability because IQ and SAT scores were not available for the nondisabled students.

Table 1
Accuracy and Response Time Performance of RD, LD, and Nondisabled Students on Elementary-Level CAAS Tasks

Task	Group		
	Nondisabled	RD	LD
Simple accuracy[a]	98.0	97.2	94.8
	(5.05)	(9.24)	(5.55)
Simple RT[b]	.633	.725	.706
	(.164)	(.185)	(.263)
Letter accuracy	97.8	98.2	94.9
	(2.75)	(1.43)	(5.62)
Letter RT	.792	.839	.931
	(.182)	(.174)	(.178)
Word accuracy	99.1	98.5	99.3
	(1.73)	(2.61)	(0.92)
Word RT	.513	.557	.605
	(.058)	(.123)	(.174)
Pseudoword accuracy	92.2	86.8	89.0
	(6.86)	(9.51)	(8.23)
Pseudoword RT	.665	.793	.807
	(.104)	(.336)	(.186)
Category accuracy	96.1	97.4	94.4
	(3.91)	(2.85)	(8.17)
Category RT	1.25	1.31	1.47
	(.226)	(.233)	(.499)
Semantics accuracy	96.3	94.4	91.5
	(5.11)	(6.35)	(7.84)
Semantics RT	2.41	2.82	2.95
	(.480)	(.658)	(.686)

Note. Standard deviations in parentheses. RT = response time; RD = reading disabled; LD = learning disabled; CAAS = computer-based academic assessment system. $N = 68$.
[a]Accuracy measured as percent correct. [b]Response time in seconds.

result is a Z score indication of where the average RD or LD subject would score if he or she were in the nondisabled group. For further clarity of presentation, the effect sizes for the RD and LD groups were transformed into the percentiles shown in Figure 1, where the solid line at the 50th percentile represents performance of the nondisabled group (a consequence of the effect size calculation). The resulting pattern of performance of the RD and LD groups in Figure 1 is discussed in more detail in the next section.

It is worthwhile to mention why the response time data were converted into effect sizes. First, presenting the results in this way eliminates scale differences between the tasks. Second, the effect size calculation is a useful way of displaying the differential patterns of performance of the diagnostic groups, especially when what is of particular interest is how the performance patterns of disabled groups (i.e., RD and LD groups) differ from nondisabled.

Discussion

The CAAS system appeared to be successful, not only in distinguishing students with learning disabilities from those who did have a learning disability, but also in differentiating among students with different types of disabilities. Support for these conclusions can be seen in Figure 1. First, the performance of the RD and LD groups clearly differs from that of the nondisabled group. Moreover, the RD and LD groups exhibited distinct CAAS profiles. The LD group was generally slow across all tasks compared to nondisabled students, even on the simple task that has nothing to do with reading. The RD group, by contrast, exhibited a pattern where performance is close to that of the nondisabled group on some tasks, but much poorer on others. In particular, the RD group was much slower than nondisabled students on the word- and pseudoword-naming tasks, but performed comparably to the nondisabled students on the letter and category match tasks.

One unexpected result was the poor performance of the RD group on the simple task (29th percentile). Given that the task contains nonverbal stimuli and is not related to reading, performance of the RD group should be similar to nondisabled peers. On further examination of the data, the slow performance of the RD group appeared to be due to two outliers. Two RD participants had response times that were

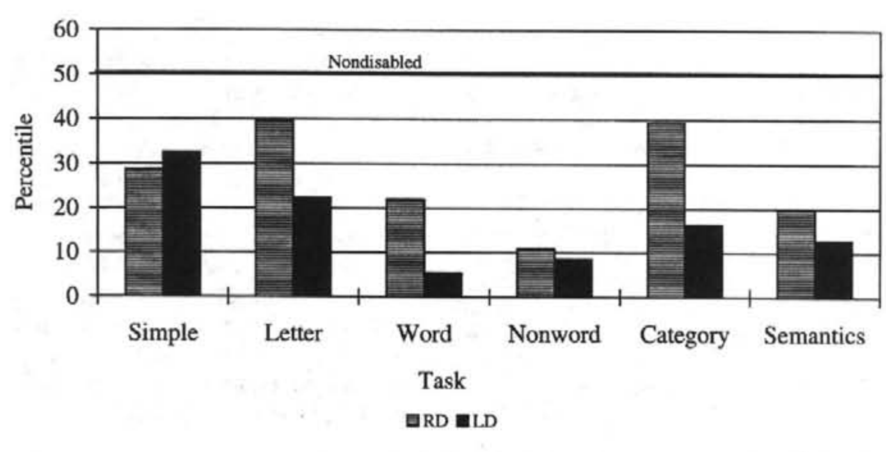


Figure 1. Percentile performance of reading-disabled (RD) and learning-disabled (LD) college students on elementary-level computer-based academic assessment system tasks as compared to nondisabled college students. (Nondisabled students are represented by the solid line at the 50th percentile.)

Table 2
Characteristics of Participants in Experiment 2

Group	n	Mean age	Gender (% female)	Ethnicity (% Caucasian)	Mean full-scale IQ
Nondisabled	42	21	67	88	
Learning disabled	34	22	62	94	101
RD	7	22	43	100	109 (99–115)[a]
LD	10	22	90	80	95 (77–103)
Other	17	21	53	100	100 (87–127)

Note. RD = reading disabled; LD = learning disabled.
[a]Range in parentheses.

over two standard deviations above (slower than) the nondisabled mean on the simple task, but these participants had response times that were within two standard deviations of the nondisabled mean on all reading tasks.

Experiment 2

The encouraging findings from Experiment 1 prompted the development of a reading battery analogous to the original elementary-level tasks but that would be more appropriate for college students (i.e., the adult-level battery). In Experiment 2, we examined whether the adult-level battery could also differentiate reading disabled, nondisabled, and other disabilities.

Method

The process of developing adult-level CAAS versions of the word, pseudoword, category and semantics tasks was relatively straightforward. A second change involved developing phonological processing tasks (described in an earlier section) to be used within the computer-based environment of the CAAS system as an attempt to measure the phonological processing problem suggested by reading disability theory and research as being at the core of specific reading disability.

Another change in CAAS tasks administered in Experiment 2 involved using naming responses rather than the button-press responses from Experiment 1. This was motivated by research indicating that the time to press a button in a decision task includes both a component associated with the participants' perceptual–cognitive discrimination and a component associated with making a decision about which button to press (e.g., Balota & Chumbley, 1984; Chumbley & Balota, 1984; Seidenberg, Waters, Sanders, & Langer, 1984). These times can be independent from one another, so a participant who was fast on the cognitive discrimination but slow on response decision could have the same response time as a participant who was slow on the cognitive discrimination but fast on the response decision. Naming tasks appear to have fewer problems of this type, so all of the CAAS tasks were converted to naming tasks.

The final change in the Experiment 2 CAAS tasks involved using a letter-naming task rather than the Posner letter-match task. The reason for this change was the notion that simple letter recognition might be more relevant to reading assessment than a task that required examinees to make decisions about whether pairs of letters had identical names.

Participants. A different group of 37 students was recruited from DSS at the same college as in Experiment 1, and a group of 42 nondisabled students who reported having no reading difficulties were recruited from introductory psychology courses at the same college. Once data collection was completed, students in the learning disability sample were classified into one of three diagnostic categories (described below): reading disability (RD), generalized learning disability (LD), or a specific learning disability other than reading (other). The demographic information for these groups as well as for the nondisabled group is shown in Table 2 and a complete description of the classification system is as follows.

Classification of learning-disabled participants into the RD, LD, and other diagnostic categories was based primarily on the clinical judgment of the counselor at DSS. The counselor was asked to describe what she thought was the student's primary difficulty based on her work with the students, which included tutoring, arranging tutors, and suggesting modifications of the curriculum to accommodate their disabilities. Other sources of data were used to supplement the counselor's clinical judgment: (a) students' self-report of their academic problems from a brief interview, (b) standardized IQ and achievement test scores,[3] (c) descriptions of students' strengths and weaknesses from evaluation report summaries, and (d) whether there was a reported history of a learning disability. The main reason for using the counselor's observations as the primary source of data for classifying students and for using the other sources as supplements was that the counselor was the only data source that was consistently available for all students. For instance, standardized test score information was missing (either missing IQ scores, achievement scores, or both) for 49% of the participants, and information about the historical nature (e.g., when and how) of previous diagnoses was missing in 65% of the cases.

On the basis of all available information in the students' records from the sources described above, Cheryl A. Cisero and a graduate student independently classified learning-disabled participants into RD, LD, or other diagnostic categories. Participants were classified as belonging to a diagnostic category if information from all available sources was consistent in indicating a reading disability, generalized learning disability, or other disability. A participant was not classified whenever information from different sources was inconsistent or when information was ambiguous (as was the case with 3 participants). As a result, the RD group contained students

[3] A student would be considered reading disabled if he or she had a Full-Scale IQ, Verbal IQ, or Performance IQ score (any of these were used as sometimes only one of them was provided) of at least 85 and standardized reading scores at least 2 years below grade level or below the 30th percentile. Because students were required to obtain diagnostic evaluations from outside professionals, there was considerable variability in the types of standardized tests used and the types of scores reported (as well as much missing information), which precluded any systematic classification of students into diagnostic categories based on test scores.

who exhibited problems in areas such as reading comprehension, word recognition, or decoding, or who were considered slow readers. The LD category consisted of students who exhibited deficits in multiple academic or cognitive domains (reading, mathematics, written language, auditory processing, etc.). The other diagnostic category contained students who had specific disabilities in a domain other than reading (e.g., math, written language, attention deficit disorder). Given the high agreement between the Cheryl A. Cisero and the graduate student rater in classifying disabled students (90.9%), only the classifications of Cisero were used in analyses.

Tasks and procedure. Participants were individually administered the elementary-level and adult-level CAAS tasks that were previously described. The elementary-level tasks were identical to those used in Experiment 1 except for replacing the letter-match task with the letter-naming task and replacing button-press responses with naming responses. All adult-level tasks were also naming tasks.

Tasks were presented in the following sequence: simple response time task, letter naming, word naming (elementary- then adult-level), pseudoword naming (elementary then adult), phonological processing tasks (rhyme, initial phoneme, final phoneme), category match (elementary, then adult), and semantics (elementary, then adult) tasks. The adult-level tasks were presented after the corresponding elementary-level tasks to decrease the amount of time needed for instruction because the task requirements for elementary and adult versions were similar.

For each task, participants were told to respond as quickly and as accurately as possible. Participants were also given practice trials and were allowed to ask questions prior to each task. Administration of the elementary and adult tasks took approximately 1 hr.

Results

Experiment 2 had two main goals: (a) to attempt to replicate the findings of Experiment 1 using the elementary-level CAAS reading battery, and (b) to determine whether similar results could be obtained when using the adult-level CAAS reading battery. Therefore, the results regarding whether disabled and nondisabled groups of college students could be differentiated by their CAAS performance are discussed separately for the elementary-level battery and the adult-level battery. The elementary battery consisted of the elementary versions of the word, pseudoword, category, and semantics tasks, and the adult battery consisted of the adult versions of the word, pseudoword, category, and semantics tasks, and a composite measure of the three phonological processing tasks. Because the simple and letter tasks are considered a part of any grade-level CAAS battery (i.e., elementary, middle school, adult) that is administered, these tasks were included in analyses of both the elementary and adult batteries. Table 3 displays the accuracy and response time performance of students in the diagnostic categories (RD, LD, other, nondisabled) on all CAAS tasks.

Elementary-Level CAAS reading battery. As in Experiment 1, Kruskal-Wallis tests were performed on the accuracy data and the response time data were subjected to a multivariate analysis of variance with diagnostic category (RD, LD, other, nondisabled) as a between-subject factor and task as a within-subject factor. With respect to accuracy,

Table 3
Accuracy and Response Time Performance on Elementary-Level and Adult-Level CAAS Tasks of Students in Different Diagnostic Categories

	Diagnostic category			
Task	Nondisabled (n = 42)	RD (n = 7)	LD (n = 10)[3]	Other (n = 17)
Simple accuracy[a]	98.9 (2.59)	100.0 (0.0)	97.8 (5.08)	99.0 (1.73)
Simple RT[b]	.569 (.121)	.578 (.109)	.819 (.344)	.604 (.092)
Letter accuracy	99.5 (1.68)	99.2 (2.22)	100.0 (0.0)	99.7 (1.43)
Letter RT	.526 (.083)	.524 (.076)	.624 (.063)	.549 (.090)
Elementary word accuracy	99.2 (1.64)	99.5 (1.30)	98.9 (2.23)	99.4 (1.29)
Elementary word RT	.537 (.088)	.633 (.094)	.685 (.076)	.560 (.082)
Elementary pseudoword accuracy	94.5 (5.92)	85.6 (8.78)	90.9 (6.32)	95.2 (4.49)
Elementary pseudoword RT	.790 (.452)	1.23 (.433)	1.81 (1.45)	.846 (.282)
Elementary category accuracy	97.7 (4.22)	99.2 (1.99)	96.4 (3.43)	98.2 (2.51)
Elementary category RT	1.18 (.225)	1.52 (.338)	1.75 (.473)	1.31 (.280)
Elementary semantics accuracy	96.8 (6.21)	98.4 (2.79)	98.3 (2.80)	96.7 (5.25)
Elementary semantics RT	1.74 (.330)	2.35 (.484)	2.67 (.725)	1.99 (.422)
Adult word accuracy	93.3 (5.78)	86.7 (6.34)	85.1 (7.41)	93.3 (4.65)
Adult word RT	.729 (.248)	.966 (.168)	1.47 (.616)	.765 (.191)
Adult pseudoword accuracy	87.8 (10.44)	78.9 (16.56)	81.7 (9.94)	91.1 (6.79)
Adult pseudoword RT	1.16 (.791)	2.09 (.670)	2.20 (.764)	1.43 (.738)
Adult category accuracy	90.5 (8.57)	88.3 (6.26)	90.9 (9.23)	88.5 (8.01)
Adult category RT	1.59 (.358)	2.22 (.579)	2.79 (.755)	1.81 (.471)
Adult semantics accuracy	89.7 (10.13)	90.6 (9.07)	85.6 (7.97)	89.9 (9.61)
Adult semantics RT	2.86 (.732)	3.53 (.950)	5.01 (1.05)	3.28 (.942)
Adult phonological accuracy	89.3 (5.65)	84.5 (5.83)	88.8 (4.10)	90.4 (3.78)
Adult phonological RT	1.49 (.476)	2.20 (.543)	3.17 (1.19)	1.78 (.502)

Note. Standard deviations are in parentheses. RD = reading disabled; LD = learning disabled; RT = response time; CAAS = computer-based academic assessment system.
[a]Accuracy measured as percent correct. [b]Response time in seconds.

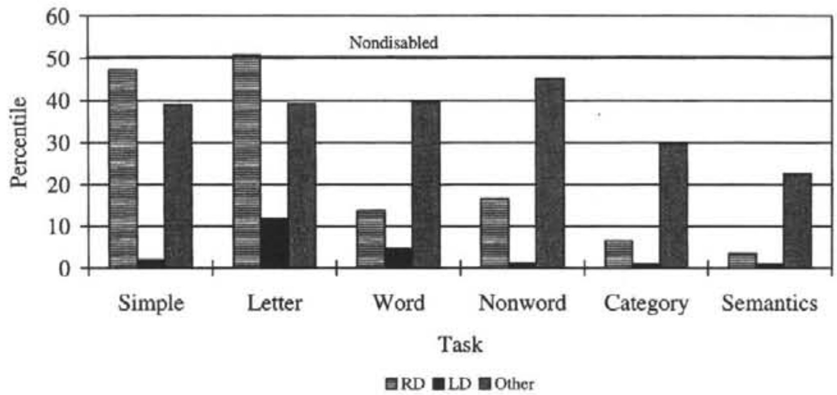


Figure 2. Percentile performance of college students from reading-disabled (RD), and learning-disabled (LD), and other diagnostic groups on elementary-level computer-based academic assessment system tasks as compared to nondisabled college students. (Nondisabled students are represented by the solid line at the 50th percentile.)

the diagnostic groups were found to be significantly different only on the pseudoword task, $\chi^2(3, N = 76) = 13.85$, $p < .01$, as in Experiment 1. Also similar to Experiment 1, the response time analysis resulted in a significant effect of task, $F(5, 360) = 228.92$, $MSE = .09$, $p < .001$, where response time increased as a function of task complexity, and a significant effect of diagnostic category, $F(3, 72) = 15.46$, $MSE = .28$, $p < .001$. A set of planned contrasts comparing RD, LD, and other groups to the nondisabled group (to control for Type I error, the Bonferroni inequality was used to set alpha at .017) indicated that RD and LD groups were significantly slower overall than the nondisabled group, RD, $t(72) = -2.89$, $SE = .217$; LD, $t(72) = -6.56$, $SE = .187$.

There was also a significant interaction between diagnostic category and task for the response time data, $F(15, 360) = 5.89$, $MSE = .09$, $p < .001$. The nature of this interaction was that the pattern of performance across the CAAS tasks differed for the diagnostic groups, as in Experiment 1. As displayed in Figure 2 (again performance is shown as percentiles converted from effect sizes), the response time performance of the other group was similar in general to the nondisabled group; the RD group was also comparable to the nondisabled group in response time performance on the simple and letter tasks, but much slower on the remaining tasks; and the LD group appeared to be considerably slow on all tasks.

Adult-level CAAS reading battery. Again, accuracy data were examined using Kruskal-Wallis tests and response time data were evaluated in a multivariate analysis of variance with diagnostic category[4] (RD, LD, other, nondisabled) as a between-subject factor and task as a within-subject factor. A significant difference in accuracy performance among diagnostic groups was found on the word and pseudoword tasks: word, $\chi^2(3, N = 76) = 14.98$, $p < .01$; pseudoword, $\chi^2(3, N = 76) = 8.92$, $p < .05$, where the RD and LD groups were less accurate than the nondisabled and other groups on these tasks. With respect to the response time analysis, there was a significant effect of task, $F(6, 426) = 292.99$, $MSE = .19$, $p < .001$, and a reliable effect of diagnostic category, $F(3, 71) = 22.78$, $MSE = .84$, $p < .001$, as was found with the elementary-level analysis. Planned contrasts on the response time data (using alpha at .017 to control for Type I error) indicated that the RD and LD groups were significantly slower overall than the nondisabled group: RD, $t(71) = -3.21$, $SE = .374$; LD, $t(71) = -8.02$, $SE = .337$, as in the elementary-level data.

More interesting, however, is the interaction of diagnostic category and task obtained for the response time data, $F(18, 426) = 7.80$, $MSE = .19$, $p < .001$. The interaction is shown in Figure 3 as percentile performance (converted from effect sizes on the response time scores) of the disabled groups relative to the nondisabled group. Figure 3 shows a pattern where the LD group is very slow on all CAAS tasks relative to nondisabled, whereas the RD group is slower than nondisabled on the adult-level reading tasks but not on the simple and letter tasks. Performance of the other group, also shown in Figure 3, is quite similar to that of the nondisabled group, except for a drop in percentile scores (i.e., slower response times) on the category, semantics, and phonological tasks. One possibility for this drop in performance is that the group, which contained several students with attention deficit disorder, may have been fatiguing on the tasks that were presented at the end of the battery.

Discussion

Experiment 2 replicated the results of Experiment 1 in that the elementary CAAS battery differentiated between students with disabilities and those who did not have a disability and distinguished among students with different kinds of learning disabilities. Experiment 2 also provided evidence that the newly developed adult version of the CAAS system was also capable of making these distinctions.

[4] One student in the LD category was excluded from the analysis of the adult-level data because of missing data on the phonological processing tasks.

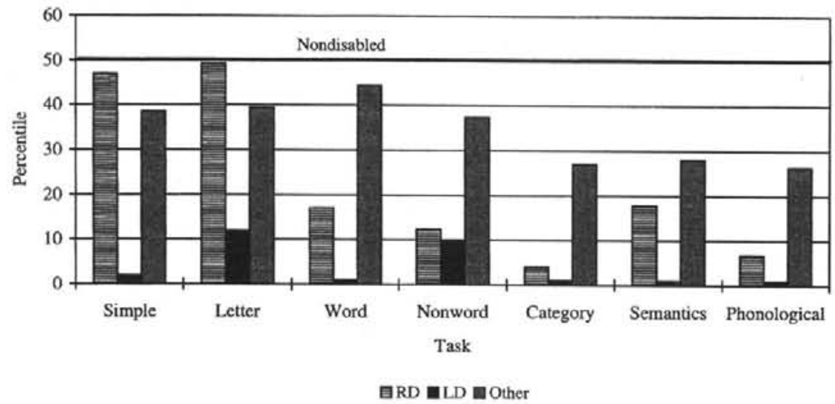


Figure 3. Percentile performance of college students from reading-disabled (RD), learning-disabled (LD), and other diagnostic groups on adult-level computer-based academic assessment system tasks as compared to nondisabled college students. (Nondisabled students are represented by the solid line at the 50th percentile.)

Several aspects of the results support these conclusions. First, the RD and LD groups were found to be significantly slower on the CAAS elementary-level battery than the nondisabled group, as in Experiment 1. Second, different patterns of response time performance for the diagnostic groups were found on the elementary-level tasks, and even more striking, the patterns shown by the RD and LD groups were quite similar to those obtained in Experiment 1. Third, the adult-level CAAS assessment also revealed differential patterns of response time performance for the diagnostic groups that were very similar to the patterns obtained with the elementary battery. For both the elementary-level and adult-level batteries, the RD group exhibited response time performance comparable to the nondisabled students on the simple and letter tasks, but slower performance at the word identification level and beyond (a pattern that would be expected with a specific reading disability), and the LD group was extremely slow across all CAAS tasks. The results from both studies, therefore, were remarkably consistent given the different samples of disabled and nondisabled students, the different methods of classifying disabled students into groups, and the slight variations in the types of CAAS tasks used (e.g., button press in Experiment 1 and vocal responses in Experiment 2).

Having provided evidence that CAAS assessments can differentiate between students who are disabled and nondisabled and can differentiate among disabled students with different diagnoses, our research effort turned to the issue of whether CAAS profiles of reading-disabled students provide information about the specific nature of the student's disability.

Experiment 3

The first and most basic requirement of an assessment technique for diagnosing specific reading disability is that it differentiate disabled readers from nondisabled readers and from students with other learning problems. However, once a student is identified as having a reading disability, the assessment tool should also be able to provide information about the nature of the reading problem that would aid a diagnostician in developing individualized interventions.

We address this issue in Experiment 3 by exploring whether the CAAS system could reveal something about the specific nature of a student's reading problem. Experiment 3 also examines whether CAAS is a valid means of assessing reading disability using a methodology that is the inverse of the methodology used in Experiments 1 and 2. In Experiments 1 and 2, students from several diagnostic groups were administered the CAAS battery, and the data were then examined to see if the groups displayed different patterns of performance. In contrast, the methodology in Experiment 3 involved first administering CAAS assessments to students receiving services from a learning disability center, and then sorting students on the basis of their CAAS profiles. The diagnostic records of the students were then examined to see if students' CAAS profiles matched with information contained in their records.

Method

Participants. Participants were 8 students identified as reading disabled (RD) by the director of DSS at the same college as in Experiments 1 and 2. As in Experiment 1, the director identified these students as RD based on his clinical judgment of a student's disability from interviews, history, test scores, and experience with the students. Table 4 presents available demographic data for each RD subject.

Thirty-five nondisabled college students from an introductory psychology course at that college who reported no reading difficulties completed the CAAS tasks and served as a comparison group. Nondisabled students were similar to the RD group in gender composition (68% female as compared to 75% in the RD group). However, no other demographic information was obtained from the nondisabled group. It should also be noted that nondisabled students provided data used to generate comparison graphs, but played no other part in the results.

Tasks and procedure. Participants were administered the elementary-level and adult-level reading tasks used in Experiment 2 (except the phonological processing tasks, which had not yet

Table 4
Characteristics of Reading-Disabled Participants in Experiment 3

Participant	Gender	Age	IQ[a]	Grade first tested	Initial problem	Remediation	Most recent testing shows poor:
1	F	19	90	2	language, reading	language, reading	reading, mathematics, written language[b]
2	F	20	—	3	reading comprehension	none	(no recent testing)
3	F	20	94	4	math	math	(average reading, math, written language[c])
4	F	19	102	5	reading, writing	reading, writing	written language[e]
5	F	19	113	5	reading	reading	reading, spelling,[c] reading rate[d]
6	M	19	89	8	visual sequencing	none	reading comprehension, written language[b]
7	M	20	112	college			reading rate, auditory memory[e]
8	F	19	139	college			reading comprehension, applied mathematics[b]

Note. F = female; M = male.
[a]Full Scale IQ from Wechsler Adult Intelligence Scale-Revised, except for Participant 6. [b]Woodcock-Johnson Psycho-Educational Battery. [c]Wide Range Achievement Test—Revised. [d]Nelson Denny Reading Test. [e]No tests reported.

been developed at the time of this research). Tasks were administered in the following order: simple task, letter task, word naming (elementary- then adult-level), pseudoword naming (elementary then adult), category match (elementary, then adult), and semantics (elementary, then adult). The entire battery took about 45 min to administer.

Results and Discussion

A CAAS profile of performance similar to the group performance shown in Figures 1–3 was constructed for each student. However, rather than constructing a separate accuracy and response time profile for each student, accuracy and response time data on each task were combined to yield a single overall profile based on both measures. A transformation procedure developed by Sinatra and Royer (1995) was used to combine a student's accuracy and response time performance on each task into a single index.[5] Each RD student's combined index on each task was then converted into an effect size (in the same way that effect sizes were calculated for response time group data in Experiments 1 and 2). To simplify the presentation of data, the effect sizes of RD students were transformed into percentiles (where the solid line at the 50th percentile represents performance of the nondisabled group).

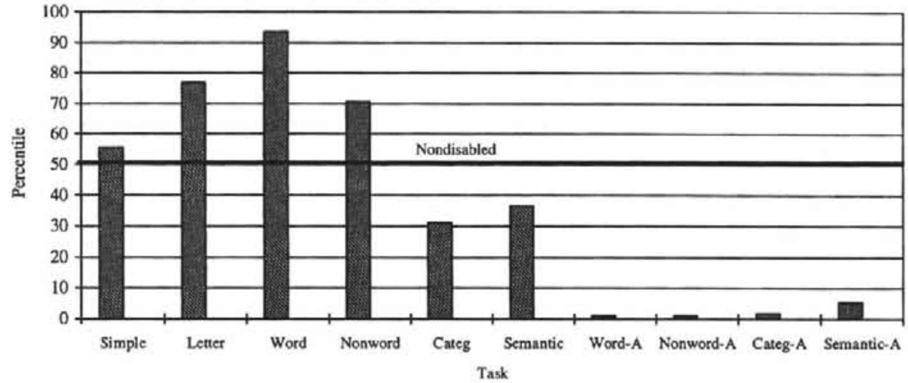
The CAAS profiles of RD participants were then examined to determine whether there were similarities in their patterns of performance. Four different profile types emerged: a "compensatory reading disability" profile (n = 2), a "severe reading disability" profile (n = 2), a "non-reading-disabled" profile (n = 3), and a "variable" profile for one student who did not appear to fit any of the other profile types. Figure 4a–d displays the CAAS percentile performance of RD participants who are representative of the four different profiles of performance. Following division of the students into profile types, their diagnostic records were examined to determine if aspects of the diagnostic record fit the patterns in their CAAS profiles.

The profile in Figure 4a is of a student who represents the compensatory reading disability profile. This profile reveals performance that is comparable to nondisabled peers on the simple and letter tasks and on most elementary tasks, but performance that is substantially poorer on all adult tasks. This student's performance on the adult tasks is consistent with documentation of below-average achievement for her age on reading subtests of the Woodcock-Johnson Psycho-Educational Battery. The fact that the student performs well on elementary CAAS tasks suggests that she has learned to compensate to some extent for her disability. Her ability to compensate on relatively simple material may be due to receiving remedial instruction for her disability since Grade 2. Similarly, the other student showing this profile had also received remediation for her reading difficulties (as well as for writing) since elementary school. The profile shown in Figure 4a suggests that students who show a compensatory pattern would benefit most from an intervention that focuses on improving word identification skills using adult-level words.

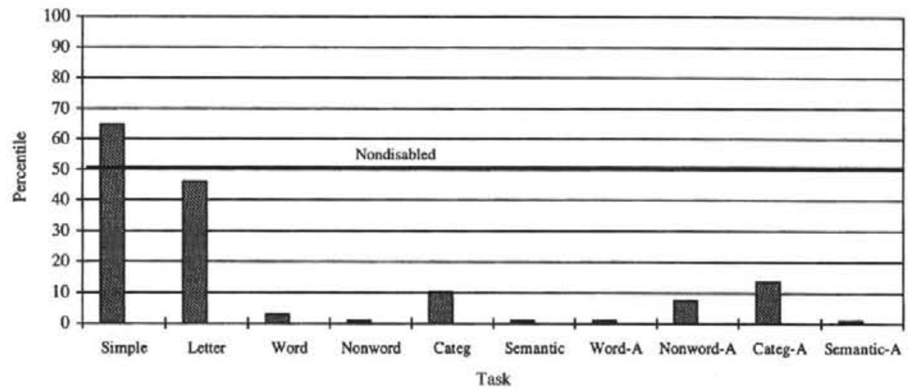
The pattern shown in Figure 4b is of a student who represents the severe reading disability profile. This student performs at the level of nondisabled peers on the simple and letter tasks. However, her performance on all other elementary and adult reading tasks is at or below the 10th percentile. This is consistent with standardized test scores that indicate below-average word reading, spelling, and reading rate. Like the student in Figure 4a, this student was identified as reading disabled in elementary school and

[5] The procedure for combining response time and accuracy into a single index was developed midway through the research effort reported in this article. The impetus for developing the combined index was provided by research with children where it was sometimes the case that students were very fast but very inaccurate (i.e., displayed a speed–accuracy trade-off). The technique developed by Sinatra and Royer (1995) combines response time and accuracy into an index that varies in accordance with expert judgments about reading skill differences between students. The new index was used in Experiment 3 because it provided a convenient means of summarizing patterns of accuracy and response time data. The reader is directed to Sinatra and Royer (1995) for details on calculating this index.

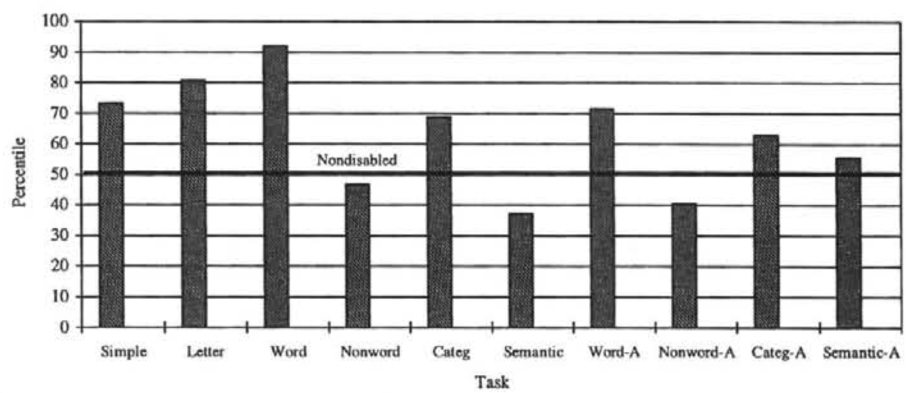
A


Percentile
100
90
80
70
60
50
40
30
20
10
0
Nondisabled
Simple
Letter
Word
Nonword
Categ
Semantic
Word-A
Nonword-A
Categ-A
Semantic-A
Task

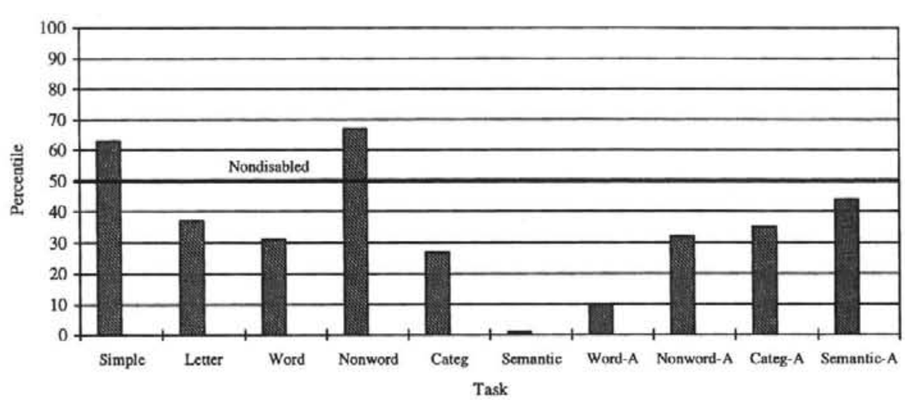
B


Percentile
100
90
80
70
60
50
40
30
20
10
0
Nondisabled
Simple
Letter
Word
Nonword
Categ
Semantic
Word-A
Nonword-A
Categ-A
Semantic-A
Task

C


Percentile
100
90
80
70
60
50
40
30
20
10
0
Nondisabled
Simple
Letter
Word
Nonword
Categ
Semantic
Word-A
Nonword-A
Categ-A
Semantic-A
Task

D


Percentile
100
90
80
70
60
50
40
30
20
10
0
Nondisabled
Simple
Letter
Word
Nonword
Categ
Semantic
Word-A
Nonword-A
Categ-A
Semantic-A
Task

subsequently received remediation. However, her CAAS profile indicates that her reading skills, even with respect to familiar vocabulary, are still severely impaired. Hence the name "severe reading disability" profile. This type of profile suggests an intervention that would focus on word identification skills, as does the compensatory RD profile, except that training would begin with words at a much lower level than for students showing the compensatory pattern.

The profile displayed in Figure 4c represents a non-reading-disabled profile. This student's performance is comparable to nondisabled peers on almost all elementary and adult tasks. Her nondisabled CAAS performance is consistent with the fact that she exhibits above-average reading skills on the Letter-Word Identification and Word Attack subtests of the Woodcock-Johnson Psycho-Educational Battery but experiences difficulty in other areas not measured by CAAS such as reading comprehension and applied mathematics. The other 2 students who showed a nonreading-disabled CAAS profile also had documentation in their records suggesting relatively normal reading skills. One student had been receiving services for a mathematics disability since the fourth grade, and had age-appropriate standardized reading test scores at her most recent testing. The other student had reading comprehension difficulties in Grade 3 but no updated testing to determine any current reading problems. Thus, although these students had been identified as reading disabled through conventional methods, their diagnostic records, as well as their CAAS profile, suggested normal reading skills.

With respect to intervention, therefore, the students included in Figure 4C are the easiest to deal with from our perspective. They might very well benefit from study strategy interventions (as would many other college students), but their inclusion with reading-disabled students is a clear case of erroneous classification that can result from traditional methods of identifying learning disabilities.

The final profile, "variable," exhibited by a single student is shown in Figure 4d. This profile is called "variable" because the student's performance appears to fluctuate according to the order of presentation of the tasks. At the beginning on the simple task, performance is comparable to nondisabled performance. The student's performance begins to drop on the letter-naming and elementary- and adult-level word-naming tasks, which are at about the middle of the test battery. Performance picks up again on the elementary-level pseudoword-naming task and then drops off on the adult-level pseudoword task, the category match tasks, and the elementary-level semantics task. The student's performance then appears to rebound on the last task, adult-level semantics. It is interesting to note that like the students who exhibited the nonreading-disabled profile, this student's standardized test scores do not indicate any specific reading disability. It is not possible to make further interpretations of this CAAS profile without any other information regarding the student's academic difficulties. In fact, when this research was first conducted, this profile seemed to be an anomaly that did not make sense. However, we have since seen this type of profile in more recent research, albeit rarely, and the profile is typically exhibited by students who have been diagnosed with attention deficit disorder (e.g., Cisero, 1996; Royer, 1996).

The CAAS profile shown in Figure 4d shows no obvious place for targeting interventions. It is quite likely that the student has problems in reading comprehension, but the CAAS profile suggests that the problems may not be due to any problem that the CAAS system can identify. It is possible that an effective intervention for this type of student may be to concentrate on training attentional processes rather than on developing reading skills.

In sum, students' performance on CAAS tasks was generally consistent with history information and standardized test performance. CAAS assessments, though, provide much more detailed information regarding the nature of the student's reading difficulty than do standardized test scores. To illustrate, the students representing the compensatory and severe reading disability profiles (Figure 4a and b) both had below-average performance on word recognition subtests of standardized tests but exhibited quite distinct CAAS profiles indicating somewhat different reading problems. Therefore, in contrast to single test scores, the CAAS profile provides the diagnostician with an integrated "picture" of a student's disability that could be valuable in planning remediation. The next study presents evidence that information obtained from the CAAS technique, such as that described here, can be used to develop interventions that remediate the reading impairments.

Experiment 4

The Laboratory for the Assessment and Training of Academic Skills (LATAS) at the University of Massachusetts provides assessment services and develops and implements educational intervention plans for students experiencing difficulties in reading, math, and writing. Most of the students receiving LATAS services are in elementary or secondary school, though occasionally adults are referred. CAAS assessments are used at LATAS to provide information about the nature of the student's difficulty, to target academic skills for educational intervention, and to monitor the effectiveness of intervention efforts.

When a student is referred to LATAS, they complete a listening and reading comprehension test (based on the

Figure 4. (opposite). Percentile performance of a reading-disabled college student from the compensatory reading disability (A), severe reading disability (B), non-reading-disabled (C), and variable (D) profile categories of the computer-based academic assessment system. (Nondisabled performance is indicated by the solid line at the 50th percentile.) Categ = category.

Sentence Verification Technique procedure; for a description, see Royer, Carlo, & Cisero, 1992) and the CAAS battery in both reading and math. Typically, the student takes an age appropriate test (e.g., elementary CAAS battery if the student is in elementary grades), but sometimes a student is administered a lower-level battery if the student's case history presents a particularly severe disability. So, for example, a Grade 7 student might take the elementary CAAS battery (suitable for students in Grades 1–4) rather than the middle school battery if the record suggests severe dyslexia.

Once the assessment process is complete, an intervention plan is developed that targets the lowest level skill identified by the CAAS assessment that is giving the student problems. If a student has a reading difficulty, the problem skill is typically word identification, though LATAS has also seen students who have difficulty at the letter identification level, and in several cases, students who have perfectly normal CAAS profiles except for very poor performance on the semantics task (and on the reading comprehension test). In the latter case, the intervention approach focuses on reading comprehension rather than lower level reading skills.

Below are case studies of 2 students who have received services at LATAS as evidence that the CAAS system can provide prescriptives that lead to effective interventions.

Case Study of G.C. (8th Grade Boy)

G.C. (8th-grade boy) came to LATAS in the summer of 1994, shortly before he started the 8th grade. G.C. was referred to LATAS by a clinical psychologist who specialized in working with learning-disabled adolescents. G.C. had been diagnosed as having attention deficit with hyperactivity disorder (ADHD) and was medicated during school hours. The fact that G.C. had ADHD and was taking medication for the disorder added variability to our assessment process. There were times when he would arrive at LATAS still under the influence of his medication, and his performance at those times would be strikingly different from when he arrived after his medication had worn off. During the time G.C. was receiving services at LATAS, he and his family were also receiving psychological services aimed at improving his behavior at home and at school.

G.C.'s academic performance in early elementary school was about average, but during his Grade 5 school year his performance began to deteriorate markedly. G.C. was then formally evaluated by his school system and began to receive special education services in reading in Grade 5. His average Grade 7 grades were in the D range, and he received four Fs in core academic subjects during the four marking periods of the Grade 7 school year.

G.C.'s performance on the reading sections of standardized tests was consistently low, though somewhat variable. For instance, at the beginning of Grade 7 he scored at the 34th percentile, and at the beginning of Grade 8, at the 4th percentile. His math, social studies, and science performance on the standardized tests was in the average range.

CAAS performance profile. On his first visit, G.C. was administered the elementary version of the CAAS battery (given the reports of his poor academic performance that we had prior to assessing G.C.). G.C.'s pattern of percentile performance on the CAAS tasks relative to other Grade 8 students is depicted in Figure 5 (the percentiles are derived from the composite accuracy and response time measure described in Footnote 5). G.C.'s pattern of performance is somewhat unusual in that we generally see substantial correspondence between performance on the simple and letter tasks, and G.C.'s performance on these tasks is very different. A possible reason for the difference between the tasks is that G.C.'s attentional problems often caused considerable variability both between and within assessments. What is obvious, though, in Figure 5, is that G.C.'s performance is considerably depressed on all of the CAAS reading tasks.

Because the CAAS assessments indicated that the lowest level reading skill that was giving G.C. trouble was word recognition, we began our intervention with a task designed to build speed and accuracy of word identification ability. G.C. arrived at LATAS with the impression that the reading activities he would have to complete would be trivial and boring (a belief derived from his experiences in special education classes). Accordingly, we wanted to present him with challenging exercises. At the time, we also felt that it was important that G.C. improve his ability to sound out words he did not recognize. These two considerations led to using commercially available English vocabulary words (*Vis Ed* cards) to build G.C.'s word identification skills. A box contains 1,000 words that are difficult even for college-level readers (e.g., *permeate, etymology, prevaricator, gyration, nefarious*).

G.C.'s word-naming intervention. G.C.'s intervention involved his selecting 80 words from the box in numbered sequence, dividing them into piles of 20, and pronouncing each of the 20 words with the help of his parent. He then repiled the 20 words and named them as fast as possible, with his parent recording the time. This procedure was repeated with the remaining three piles of 20 words, and G.C. then computed the mean time per pile and recorded the time on a graph. G.C. and his parent would repeat this exercise 5 nights a week, taking a new stack of 80 cards each night, thereby making one pass through the 1,000 words in

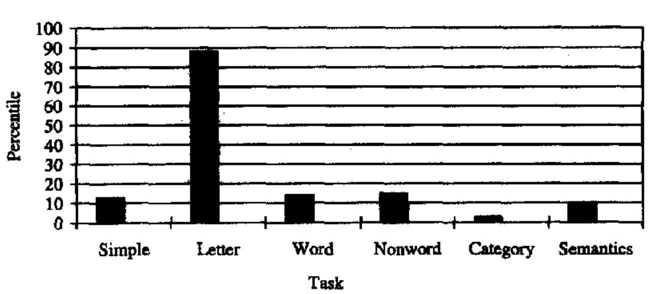


Figure 5. G.C.'s grade-level percentile performance on the Laboratory for the Assessment and Training of Academic Skills assessments of the computer-based academic assessment system.

12 nights. During this phase of the intervention no effort was made to teach G.C. the meaning of the words.

The instructions to G.C. and his parents were to try to be as accurate as possible but to guess at a word that was taking him a lot of time. His parents were instructed to make a mental note of words that he got wrong and to return to those words after the time trial and have him pronounce the word correctly. Words that G.C. repeatedly missed, even after being corrected at the end of a trial, became part of a separate practice pile that were practiced both before and after a time trial.

Word-naming intervention results recorded at home. It took G.C. an average of 47 s to name 20 words (2.35 s per word) on the first night of training and his performance on the next 8 nights of training varied around the 45-s level. On Day 10 his timed performance began a consistent decline (representing increasing word-naming speed), dropping from 43 s on Day 10 to 16 s (.8 s per word) on Day 26. During this period of precipitous drop in naming times, G.C. became very involved in his performance and around Day 20 he asked what the world record was for word naming because he thought he could beat it. Note that this improvement began before G.C. had made one complete pass through the box of words, though his speed accelerated sharply as he began encountering words for the second time. Training on word naming continued to Day 37, where his performance leveled off to about 14 s per 20 words (0.7 s per word). At the end of word-naming training G.C.'s parents reported that his accuracy was near 100%.

G.C.'s word meaning intervention. After it was obvious that G.C. had plateaued on the word-naming exercise, we began word-meaning training. This activity was identical to the word-naming training except that rather than naming the word, G.C. said anything that indicated that he knew the meaning of the word. For example, he might provide a synonym, an antonym, or word providing context (e.g., saying "lightbulb" if the word were *electricity*). Again, G.C.'s parents recorded his time on four 20-word sets of cards, and G.C. calculated the mean for 20 words and plotted it on a graph. As with word naming, G.C. practiced the word meaning activity 5 nights per week. His time trials occurred after he had the opportunity to study the meaning of the words appearing on the back of each card.

Word-meaning intervention results recorded at home. The first night that G.C. attempted the word-meaning activity, it took him an average of 150 s per 20 words (7.5 s per word). By Day 6 he had reached a level of 70 s per 20 words (3.5 s per word) and stayed around that level up to Day 46, when training was terminated. During the training period his word meaning accuracy consistently improved to the point where G.C.'s parents reported a level near 100%.

Word-naming and word-meaning training on textbook words. Following the completion of G.C.'s training with the *Vis Ed* words, he began working at home on a word list derived from his social science textbook. This list was generated by copying the words from the textbook glossary, randomly arranging 40 words per page, and having him practice naming four pages of words per night using the procedure described earlier. After his home graph leveled out, indicating that further improvement was unlikely, G.C. completed word meaning exercises on the word lists. As an aside, rather than arranging textbook words in random order, we now arrange them in chapter order so that students will master new words in the order in which they will be encountered. G.C.'s home-recorded results on the textbook words were very similar to those recorded for the *Vis Ed* cards.

Results of Repeated CAAS Assessments

During the time that G.C. was working on word-naming and word-meaning exercises, he returned to LATAS once a week and was reassessed on the CAAS system. Initially, G.C. was assessed using the elementary CAAS battery, but after 10 weeks of training he was both very fast and very accurate on the elementary battery so we switched his assessment to the middle school battery, which contains materials at the Grades 5–8 difficulty level. There are two points that the reader should keep in mind while examining results of these assessments. First, the materials on which G.C. was being trained (the *Vis Ed* and textbook glossary words) were completely different from those contained in the CAAS battery. Second, each CAAS assessment on the word task was different because on any given assessment 40 words are randomly drawn from a pool of 240 words. These two features assure that the CAAS assessments provide an index of the extent to which word-training transfers to materials *not* being practiced.

G.C.'s accuracy and response time on repeated CAAS assessments are reported in Table 5. As can be seen in the table, his performance generally improves in both accuracy and response time during the period when home training was occurring. It is also apparent, though, that he makes more progress on some tasks than others. For instance, his improvement is much more apparent on the word, category, and sentence tasks than on the nonword task.

An important aspect of the intervention involved rewarding G.C. for improved performance on the weekly CAAS retesting. LATAS uses a grab-bag technique to encourage

Table 5
G.C.'s Accuracy and Response Times (RTs) on CAAS Tasks During the Training Period

Training week	Word		Nonword		Concept		Sentence	
	%[a]	RT[b]	%	RT	%	RT	%	RT
1	96	.90	81	.92	85	1.6	90	3.0
4	97	.68	89	.81	75	1.4	60	2.3
8	100	.58	92	.72	90	1.4	91	2.0
10	90	.89	92	1.3	70	2.9	100	3.9
14	89	.62	64	.98	78	2.8	82	1.5
18	98	.64	92	1.2	72	2.1	75	2.2
22	95	.60	92	1.2	82	2.6	80	1.2
26	96	.71	85	1.2	100	2.6	93	2.2

Note. Switched to middle school tasks at Week 10. CAAS = computer-based academic assessment system.
[a]Percent correct. [b]Response time in seconds.

Table 6
G.C.'s Grades Before and After LATAS Training

Grade level and marking period	Reading	Math	Science	Social studies	English	Algebra	Science	World history
Grade 7								
1st quarter	D+	C−	D	F				
2nd quarter	C+	F	F	F				
3rd quarter	C−	C−	C+	D				
4th quarter	D−	C	C+	D+				
Final grade	D+	D+	D+	F				
Grade 8								
1st quarter	D	C	D	D−				
2nd quarter	C+	C−	C	D−				
3rd quarter	D+	C	C−	D−				
4th quarter	C	B+	C	D				
Final grade	C−	C+	C−	D−				
Grade 9								
1st quarter					B−	A−	A	B+

Note. G.C. came to LATAS in late summer between Grades 7 and 8. G.C. stopped coming to LATAS near the end of the summer between Grades 8 and 9. LATAS = Laboratory for the Assessment and Training of Academic Skills.

students to keep practicing at home. At the first meeting with students and their families, we describe the interventions and the weekly retesting and tell students that they can go into a grab bag (which also contains envelopes of money from $0.50 to $10.00) if they improve their performance on the majority of the CAAS tasks on which they are retested. Improved performance is defined as increasing accuracy or response time by a specified amount without a corresponding drop in the other measure. So, for instance, improvement in sentence performance would occur if response time decreased by 20 ms with a 1% or less drop in accuracy.

G.C.'s mother allowed us to copy his Grade 7 and Grade 8 report cards during the period he was at LATAS. After he left LATAS, his mother sent us a copy of his report card for his first Grade 9 marking period. The grades he received in Grades 7–9 are reported in Table 6.

Case Study of K.K. (10th-Grade Girl)

K.K. was referred to LATAS by her high school guidance counselor after having just finished 10th grade. K.K.'s academic difficulties began in Grade 4. Her mother reported remembering a comment of one of her 4th-grade teachers that she appeared to be "missing something in language skills." K.K.'s academic performance hovered in the low-average range during Grades 4–8, but her mother believed that her performance was actually lower than this. K.K.'s mother said that teachers always went out of their way to comment on what a delight she was to have in the classroom, and she believed that K.K.'s teachers always gave her a break on grading because she was such a nice girl.

In Grade 8, K.K.'s grades began to dip considerably, with particular difficulties in science. These difficulties continued into Grades 9 and 10. K.K. was a very talented athlete and had aspirations of going to college. She sought advice about college from her guidance counselor, who told her that her current grades indicated that she would have a very tough time in college and recommended that she try the LATAS program.

CAAS and SVT performance profile. Given that K.K. wanted to go to college, we decided to administer the adult version of the CAAS battery at her first visit to LATAS and to compare her performance to college norms. The results of this assessment appear in Figure 6 (again, the percentiles are based on a composite accuracy and response-time index) and indicate that K.K.'s performance was generally well below that attained by the typical college undergraduate enrolled at a state college in Massachusetts. Her ability to rapidly and accurately identify words and understand sentences was particularly low, though she did display unexpected competency in sounding out nonwords. Given her poor performance on the word recognition task and her college aspirations, we decided to direct our intervention at her word recognition abilities using the *Vis Ed* cards that we used with G.C.

K.K.'s word-level intervention. K.K. began working at LATAS in July of 1995 with an intervention that was identical to the one originally implemented with G.C. That

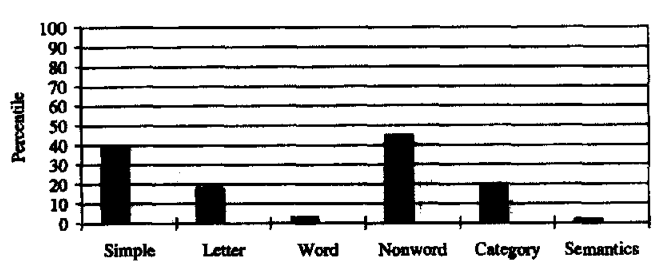


Figure 6. K.K.'s college-level percentile performance on the Laboratory for the Assessment and Training of Academic Skills assessments of the computer-based academic assessment system.

Table 7
K.K.'s Accuracy and Response Times (RTs) on CAAS Tasks During the Training Period

Training week	Word		Nonword		Concept		Sentence	
	%[a]	RT[b]	%	RT	%	RT	%	RT
1	88	.94	88	1.09	70	1.65	80	3.60
3	97	.80	90	1.04	82	1.70	89	3.60
6	89	.78	92	.91	80	1.12	89	2.70
9	91	.68	88	.82	83	1.10	89	2.26
12	97	.65						

Note. K.K. did not complete remaining tasks on Week 12. CAAS = computer-based academic assessment system.
[a]Percent correct. [b]RT in seconds.

is, 5 nights per week K.K. selected 80 words in sequence, practiced them, named 20 words at a time with her time being recorded, and then plotted the mean time to name 20 words on a graph. After completing one pass through the 1,000 words, K.K. started at the beginning and went through them again. When K.K. plateaued on word naming, she then completed the word meaning activity previously described for G.C.

Word-naming and word-meaning intervention results recorded at home. K.K.'s mean time to name 20 words was 33 s on Day 1 (1.65 s per word) and her time gradually declined until she plateaued at 20 s a word (1.1 s per word) on Day 22. Word naming training continued another 9 days and was terminated at Day 31. With respect to the word meaning activity, it took K.K. 70 s per 20 words (3.5 s per word) on the first day, and this time dropped to approximately 60 s per 20 words (3 s per word) by Day 7, where it stayed for the remainder of the word meaning training.

Word-naming and word-meaning training on textbook words. During Week 4 of the intervention, K.K. began word naming training with words from a copy of the chemistry book that she would be using in Grade 11. K.K. practiced sets of word lists (containing 4 pages of 40 words each) derived from the glossary of the textbook using the same techniques as described with G.C. These chemistry word practice sessions initially were alternated every other night with practice on the boxed vocabulary words. Shortly after beginning chemistry word training, K.K. completed her activities with the boxed words and then practiced exclusively on the chemistry words, moving from word-naming training to word-meaning training when she plateaued on the naming exercise. K.K.'s home graphs of the chemistry word-naming and word-meaning exercises looked very similar to the graphs for the boxed word naming and meaning exercises. K.K.'s Grade 11 school year started shortly after she completed the chemistry exercises and she terminated her involvement with LATAS.

Results of Repeated CAAS Assessments

K.K. was reassessed weekly at LATAS on the adult-level CAAS system. As with G.C., K.K. was allowed to go into the grab bag if her performance improved on the majority of the tasks. The CAAS retesting results are reported in Table 7. As shown in Table 5, K.K. got more accurate and faster on all of the CAAS tasks. Again, it is worth emphasizing that the improvement was on materials that she was not practicing. This suggests the progress is attributable to general gains in reading skill rather than to practice on tested materials.

Table 8 displays K.K.'s grades for Grade 10 before she came to LATAS and for her first Grade 11 marking period. The table shows that in Grade 10 she typically received the lowest grade she could get (65) without getting an F in the course. Consistent with her mother's hypothesis about teachers giving her a break on grades, three of the four Grade 10 teachers noted that K.K. was a pleasure to have in the class. K.K.'s grades in the first marking period of Grade 11 were better than those in the first marking period of Grade 10, and in a note that accompanied K.K.'s Grade 11 report card, her mother indicated that K.K. (and Mom) was particularly happy with her performance in the chemistry course.

Discussion

Experiment 4 involved two case studies that were designed to evaluate whether information provided by the CAAS system would be useful in designing effective

Table 8
K.K.'s Grades Before and After LATAS Training

Grade level and marking period	English	Geometry	Biology	History	Algebra	Chemistry	Law and Government
Grade 10							
1st quarter	71	65	70	86			
2nd quarter	70	65	65	84			
3rd quarter	59	65	62	84			
4th quarter	60	65	65	83			
Final grade	65	65	65	82			
Grade 11							
1st quarter	65				73	78	88

Note. K.K. came to LATAS in late summer between Grades 10 and 11. LATAS = Laboratory for the Assessment and Training of Academic Skills.

intervention strategies. The CAAS assessments for G.C. and K.K. indicated that both students were poor at identifying words relative to an appropriate comparison group. The interventions with G.C. and K.K. attempted to improve general word recognition skill by having the students practice speeded word identification and word meaning activities using commercially available vocabulary materials. The interventions also attempted to improve word recognition relevant to specific courses by having both students practice speeded identification and meaning activities using words drawn from course textbooks. Both students showed improvement during the interventions in general word-recognition performance, specific (course-related) word-identification performance, and general reading performance, and also showed improvement after the interventions had ended in grades in the targeted course (as well as improvement in several nontargeted courses).

Experiment 4 has all the shortcomings of case studies in general. There is no way to be certain that G.C.'s and K.K.'s improvements in reading skill and grades resulted from the interventions designed at LATAS. Rather, they could be the result of some other events that coincided with the LATAS interventions. There are a number of reasons, however, to believe that the LATAS interventions did have a positive impact on reading performance and grades.

The first reason for believing that the LATAS interventions had a positive impact is that there is a large body of theory and research that says that interventions like those at LATAS should work. A more complete exposition of this theory and research can be found in Royer and Sinatra (1994), but a brief description follows. Students with a specific reading disability very frequently are either slow, or inaccurate, or both at identifying words. Slow word identification consumes cognitive capacity necessary for comprehension processes, thereby causing the disabled reader to have poor reading comprehension. Practice at speeded word recognition results in faster and more accurate word identification, which in turn frees up cognitive capacity that can be used to enhance higher level reading processes. Thus, both G.C.'s and K.K.'s practice on rapid word identification resulted in improvements not only in word identification but also tended to improve performance on the nonword task, the concept task, the sentence task, and grades in school.

The second reason for believing the LATAS interventions resulted in improved reading performance is that the intervention with K.K. essentially served as a replication of the intervention with G.C. The students were similar in that they both had extended histories of academic difficulty, including reading difficulties, and they both experienced very similar LATAS interventions. However, they differed in gender, age, psychological profile (e.g., G.C. was ADHD and was in psychological therapy, whereas K.K. seemed to be psychologically well adjusted); went to different school systems; and their course-related word interventions were in different content areas. The fact that the same intervention produced similar positive results with both students argues for the effectiveness of the interventions. It should also be mentioned that similar results have now been obtained with younger students who have also received LATAS services (Royer, 1996). These replications, which are not described in this article because they involve work with children in the elementary grades, add to the plausibility of the interpretation that the interventions result in improvements in reading performance.

A third reason for believing that the interventions had a positive impact is that it is difficult to imagine any other event that might have produced the positive results. Both students began their LATAS activities in the summer. G.C. continued his involvement into the next school year, but K.K. terminated her involvement with LATAS shortly after school began. Both students showed improvement when school was not in session and neither was involved in an academic activity during the summer other than LATAS. It is possible, but implausible, that some other event (or probably a separate event for each student) coincided with the LATAS intervention and produced the positive outcomes.

General Discussion

In the research reported in this article, we evaluated the CAAS system as a college-level reading diagnostic against the following criteria: (a) CAAS must be valid for the purpose of identifying reading disability, (b) CAAS patterns of performance must be consistent with reading disability theory, (c) CAAS should provide information about the specific nature of the reading problem, and (d) diagnoses obtained from the assessment must lead to prescriptive procedures that alleviate the reading problem. Experiments 1, 2, and 3 reported in this article address the first three criteria, and Experiment 4 addresses the fourth criterion.

CAAS Assessments as a Valid Reading Diagnostic

The validity evidence from the studies reported in this article takes two forms: evidence that CAAS assessments can differentiate between individuals in known diagnostic categories (Criterion 1), and patterns of CAAS performance that are consistent with those predicted from theory and research on reading disability (Criterion 2).

CAAS Validity in Terms of Differentiating Between Groups

The results of Experiments 1 and 2 support the validity of the CAAS system as a diagnostic technique for identifying reading disability in that students of different diagnostic groups were differentiated on the basis of their CAAS performance. The RD group performed similarly to nondisabled students on nonreading tasks but very poorly on all reading tasks. This pattern is quite distinct from the LD pattern showing poor performance across all CAAS tasks and from the Other group's pattern showing performance that was comparable to nondisabled students on most tasks.

CAAS Patterns and Reading Disability Theory

The most popular theory of reading disability suggests that reading-disabled individuals have a phonological core deficit that inhibits their ability to learn to recognize words (e.g., Fletcher et al., 1994; Lyon, 1995; Stanovich, 1993; Stanovich & Siegel, 1994). Inaccurate and slow word recognition performance subsequently has a deleterious impact on comprehension. This theory has three implications and a corollary that are relevant to the CAAS results reported in this article. First, the theory predicts that RD individuals should have word recognition performance that is either slow or inaccurate, or both. Second, they should display deficiencies in reading processes that occur at levels higher than word identification. Third, they should display particular deficiencies on tasks that emphasize phonological processing. Finally, the corollary says that RD individuals should not display processing difficulties on tasks that do not involve reading.

To evaluate these predictions, differences between the RD and nondisabled groups on tasks in Experiments 1 and 2 will be reported in terms of effect sizes rather than in terms of statistically significant findings (which are provided in the Results sections of each of the experiments). Effect sizes are reported here to provide an index of the actual magnitude of group differences. The effect size data (note that a minus effect size reflects performance on the poor performance side of the mean) are summarized in Table 9.

With respect to word recognition performance, the effect sizes for response time on the word task indicate that nondisabled readers outperformed RD readers on words that are known by 80% of Grade 4 students. On word accuracy, the RD group performed slightly better than the nondisabled group in Experiment 2, but somewhat poorer in Experiment 1. On the adult-level word task in Experiment 2, nondisabled readers outperformed RD students in accuracy as well as response time. Thus, RD readers were slower than nondisabled readers at identifying elementary-level words, and were slower and less accurate at identifying adult-level words.

Table 9
Effect Sizes on Elementary-Level and Adult-Level CAAS Tasks From Experiments 1 and 2

Task	Accuracy	Response time
Experiment 1		
Simple	−.16	−.56
Letter	.14	−.26
Elementary word	−.34	−.75
Elementary pseudoword	−.79	−1.23
Elementary category	.33	−.26
Elementary semantics	−.37	−.85
Experiment 2		
Simple	.42	−.07
Letter	−.18	.02
Elementary word	.18	−1.09
Adult word	−1.14	−1.08
Elementary pseudoword	−.85	−.97
Adult pseudoword	−.85	−1.17
Elementary category	.35	−1.5
Adult category	−.25	−1.76
Elementary semantics	.25	−1.85
Adult semantics	−.08	−.92
Phonological processing tasks	−.85	−1.49

Note. CAAS = computer-based academic assessment system.

Regarding performance of RD and nondisabled readers on tasks above the word recognition level, we were interested in whether there were differences on the concept-activation task and the sentence-processing task, both of which require cognitive processing above the lexical access level. Inspection of effect size data from elementary-level tasks in Experiments 1 and 2 indicates that the RD group was more accurate but slower than nondisabled on concept activation. The RD group was also slower than nondisabled on sentence processing in both Experiments 1 and 2, but the RD group was less accurate than nondisabled in Experiment 1 and more accurate in Experiment 2. Moreover, the RD group was both less accurate and slower than nondisabled on adult-level concept and sentence tasks. To summarize, then, RD readers were slower than nondisabled readers on elementary-level concept activation and sentence-processing tasks, and were both slower and less accurate on age-appropriate concept and sentence tasks.

The third expectation derived from phonological core theory is that RD readers should have particular difficulties on tasks that require phonological analysis. Evidence relevant to this prediction comes from the pseudoword-naming tasks from Experiments 1 and 2 and the set of visual phonological analysis tasks used in Experiment 2. The RD group was both less accurate and slower than the nondisabled readers on the elementary-level pseudoword-naming task (Experiments 1 and 2), on the adult pseudoword task (Experiment 2), and on the set of visual phonological analysis tasks in Experiment 2.

It is worthwhile to emphasize the particular difficulty in phonological processing of the RD group relative to nondisabled. A rough index of the magnitude of the difference in phonological processing can be obtained by averaging the effect sizes across the experiments and tasks (phonological analysis and pseudoword naming). The calculation indicates that RD readers differed from nondisabled readers in phonological processing by −.83 effect size units in accuracy and by −1.22 effect size units in response time. By comparison, averaging effect sizes across experiments for word recognition yields an average effect size of −.38 in accuracy and −.97 in response time.

The final expectation is that RD and nondisabled readers will perform similarly on tasks that do not involve reading. The simple response time task is clearly one such task in the CAAS battery, and it could be argued that the letter recognition task has essentially become a perceptual task for college students. Therefore, we will look at the results across Experiments 1 and 2 on these two tasks. Table 9 shows that in Experiment 1 the RD group was slower than the nondisabled group on both the simple and letter tasks. However, the RD group was slightly less accurate than the nondisabled group on the simple task and slightly more accurate on the letter task. The RD group in Experiment 2 showed the reverse pattern of accuracy performance, being more accurate than nondisabled on the simple task and less accurate on the letter task. However, there were essentially no differences in response time between RD and nondisabled readers on the simple and letter tasks in Experiment 2. Summing over experiments and tasks, the RD group was somewhat slower (only in Experiment 1) and slightly more accurate than the nondisabled group.

Taken as a whole, the results of Experiments 1 and 2 are consistent with expectations derived from the phonological core theory of reading disability. That is, reading-disabled students are slower and less accurate than nondisabled readers on word recognition, on reading activities above the word level, and on tasks that require phonological processing, but there are relatively small differences between the groups on tasks that do not involve reading.

CAAS Assessments as Indicators of Specific Reading Problems

A valid procedure for the purpose of diagnosing reading disability should not only be able to identify a reading disabled student but should also be able to specify the nature of the reading problem so that appropriate intervention procedures can be developed to alleviate the problem. Experiments 1, 2, and 3 provided evidence that is relevant to the issue of whether CAAS assessments contain information that is useful in identifying the specific nature of the reading problem. First, the fact that CAAS assessments in Experiments 1 and 2 differentiated between students who were grouped on the basis of their learning disability is evidence that CAAS assessments can satisfy the specific nature criterion. Further support for the specific nature criterion was evident in Experiment 3 in the fact that students with a diagnosed reading disability showed distinctive CAAS profiles, which were consistent with other documented evidence relevant to the their disability, and which were suggestive of quite different intervention strategies.

The Effectiveness of CAAS-Based Intervention Efforts

In Experiment 4, we report the results of an attempt to improve the academic performance of two students, G.C. and K.K., who were identified as having reading difficulties by their case histories and by their CAAS performance. CAAS assessments provided information used to target skills for intervention, and were used to evaluate student progress during and after the training.

The initial CAAS assessments for G.C. and K.K. suggested that their reading difficulty originated at the word identification level. Intervention, therefore, began with an attempt to improve the speed and accuracy of word identification and subsequently focused on developing speed and accuracy of accessing word meanings and of word identification using words from the glossaries of textbooks. Improvements in academic performance for G.C. and K.K. were observed on their home graphs (which showed an increase in their speed of identifying words they were practicing), on CAAS assessments (which showed an improvement in general word identification performance, ability to sound out pseudowords, ability to activate the meaning of words, and ability to process sentences), and in grades that were better after the intervention than before. As noted in the discussion of Experiment 4, it cannot be said with certainty that these positive changes were due to the interventions, but there are a number of reasons (discussed earlier) for believing that the interventions had a positive influence on academic performance.

CAAS Diagnostics Relative to Traditional Procedures

The CAAS system has some advantages over traditional procedures as a college-level reading diagnostic. One advantage is that CAAS assessments provide prescriptive information that readily translates into intervention strategies. That is, a CAAS assessment shows how a student performs on a range of reading activities from low-level to high-level tasks, thereby allowing the identification of a reading skill that may be blocking attainment of reading comprehension and targeting interventions designed to repair the deficient skill. This contrasts to traditional procedures that (at best) provide a determination that the student has a learning disability, but say nothing about designing interventions that may improve the student's academic ability. This means that CAAS assessments can potentially satisfy all four of Royer and Sinatra's (1994) criteria, whereas traditional assessment procedures generally satisfy only one.

As a parenthetical comment, the lack of prescriptive information in traditional assessment procedures most likely is related to the fact that virtually all college-level interventions involve accommodating the learning environment to the student's disability. Typical recommendations include providing more time for exams, special note-taking procedures, and special exam procedures. The CAAS-based approach to working with college students has a very different message, namely that the student's learning problem is not hopeless and it is possible to improve academic functioning and allow the student to function in the normal learning environment.

A second advantage of the CAAS system is that it is inexpensive and saves time. The college at which much of the data in this article was collected does not have the in-house capability of diagnosing learning disabilities. Students seeking services at the disability center either arrive with documentation of the disability in hand or are sent to a diagnostician in private practice where they undergo an evaluation that costs between $600 and $1,200. The evaluation typically involves several hours of individual testing and development of a formal report that may take several weeks to produce. If the students do not have the money to complete this process, they are probably out of luck. As another example, a major university with which we are familiar does have an in-house capability, but the student who had not been diagnosed before undergoes over 10 hr of individual testing and pays $400 to cover part of the assessment process (the university subsidizes the remaining cost). In contrast to the expense and time of traditional procedures, CAAS assessments can be administered by nonprofessionals in approximately an hour and the results are immediately available. Even if CAAS assessments were used only to screen students trying to beat the system (e.g., avoid a foreign language requirement), it could result in a significant savings to both students and educational institutions.

In closing, we want to emphasize two points. First, what is important in the present article is the general approach of the CAAS system rather than a particular implementation of that approach. The idea providing the foundation for the CAAS system (i.e., speed and accuracy of component reading skills are important in understanding the reading process) is present in most modern theories of reading. In addition, the tasks that comprised the original CAAS system, many of which are still used, were culled from the reading research literature. We also know of at least one other commercially available computer system that is based on the same ideas and contains similar tasks (Carver & Clark, 1996), and it would be relatively easy for anyone else with a modicum of computer know-how to develop an assessment tool that does essentially the same thing as the CAAS system.

More important, the research reported in this article does not provide convincing evidence that the CAAS system satisfies Royer and Sinatra's (1994) four criteria for a reading diagnostic. We very much subscribe to Messick's (1980) view that test validation is a process of developing an ever-expanding mosaic of evidence that converges on the interpretation that the assessment process is valid for a particular purpose. The evidence reported in this article represents one tile in the mosaic; other published research (e.g., Royer, 1996; Royer & Sinatra, 1994; Sinatra & Royer, 1993) adds additional tiles, but a considerable amount of research remains to be done before one can say that CAAS assessments are valid for identifying and treating reading-disabled students.

We believe, as do many professionals in the field of learning disabilities, that proper diagnosis and treatment of learning disabilities can be attained through establishing an intimate connection between definition, assessment, and intervention. The CAAS system fits nicely within this framework. However, continued research is needed to determine the efficacy of CAAS (or other alternatives) in fulfilling this goal.

References

Adams, M. J. (1990). *Beginning to read: Thinking and learning about print.* Cambridge, MA: MIT Press.

Anderson, J. R. (1983). *The architecture of cognition.* Cambridge, MA: Harvard University Press.

Balota, D. A., & Chumbley, J. I. (1984). The locus of word-frequency effects in pronunciation of tasks: Lexical access and/or production? *Journal of Memory and Language, 24,* 89–106.

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

Carver, R. P., & Clark, S. W. (1996, April). *Investigating reading disabilities using the rauding diagnostic system.* Paper presented at the Annual Meeting of the Society for the Scientific Study of Reading, New York, NY.

Chumbley, J. I., & Balota, D. A. (1984). A word's meaning affects the decision in lexical decision. *Memory and Cognition, 12,* 590–606.

Cisero, C. A. (1996). *Are there diagnostic alternatives to the IQ-reading discrepancy?: Evaluation of assessment techniques for identifying reading disabled college students.* Unpublished doctoral dissertation, University of Massachusetts, Amherst, MA.

Dale, E., & O'Rourke, J. (1976). *The living word vocabulary: The words we know.* Field Enterprises Educational Corporation.

Evans, L. D. (1990). A conceptual overview of the regression discrepancy model for evaluating severe discrepancy between IQ and achievement scores. *Journal of Learning Disabilities, 22,* 406–412.

Fletcher, J. M., Shaywitz, S. E., Shankweiler, D. P., Katz, L., Liberman, I. Y., Stuebing, K. K., et al. (1994). Cognitive profiles of reading disability: Comparisons of discrepancy and low achievement definitions. *Journal of Educational Psychology, 86*(1), 6–23.

Fodor, J. A. (1983). *The modularity of mind.* Cambridge, MA: Harvard University Press.

Forster, K. I. (1979). Levels of processing and the structure of the language processor. In W. E. Cooper & E. Walker (Eds.), *Sentence processing: Psycholinguistic studies presented to Merrill Garrett.* Hillsdale, NJ: Erlbaum.

Francis, W. N., & Kucera, H. (1982). *Frequency analysis of English usage: Lexicon and grammar.* Boston: Houghton-Mifflin.

Liberman, I. Y., & Liberman, A. M. (1990). Whole language versus code emphasis: A critical evaluation of the underlying assumptions and their implications for reading instruction. *Annals of Dyslexia, 40,* 51–76.

Lyon, G. R. (1995). Toward a definition of dyslexia. *Annals of Dyslexia, 45,* 3–27.

Messick, S. (1980). Test validity and the ethics of assessment. *American Psychologist, 35,* 1012–1027.

Morrison, S. R., & Siegel, L. S. (1991). Learning disabilities: A critical review of definitional and assessment issues. In J. E. Obrzut & G. W. Hynd (Eds.), *Neuropsychological foundations of learning disabilities* (pp. 79–97). San Diego, CA: Academic Press.

Myers, J. L., & Well, A. D. (1991). *Research design and statistical analysis.* New York: Harper Collins.

Perfetti, C. A. (1988). Verbal efficiency in reading ability. In M. Daneman, G. E. MacKinnon, & T. Waller (Eds.), *Reading research: Advances in theory and practice* (pp. 109–143). New York: Academic Press.

Perfetti, C. A. (1992). The representation problem in reading acquisition. In P. B. Gough, L. Ehri, & R. Treiman (Eds.), *Reading acquisition* (pp. 145–174). Hillsdale, NJ: Erlbaum.

Posner, M. I., Boies, S. J., Eichelman, W. H., & Taylor, R. L. (1969). Retention of physical and name codes of single letters. *Journal of Experimental Psychology Monographs, 79,* 1–16.

Reynolds, C. R. (1981). The fallacy of "two years below grade level for age" as a diagnostic criterion for reading disorders. *Journal of School Psychology, 19*(4), 350–358.

Reynolds, C. R. (1985). Measuring the aptitude-achievement discrepancy in learning disability diagnosis. *Remedial and Special Education, 6*(5), 37–55.

Royer, J. M. (1996). A cognitive perspective on the assessment, diagnosis and remediation of reading skills. In G. Phye (Ed.), *Handbook of academic learning.* San Diego: Academic Press.

Royer, J. M., Carlo, M. S., & Cisero, C. A. (1992). School-based uses for the Sentence Verification Technique for measuring listening and reading comprehension. *Psychological Test Bulletin, 5*(1), 5–19.

Royer, J. M., & Sinatra, G. M. (1994). A cognitive theoretical approach to reading diagnostics. *Educational Psychology Review, 6*(2), 81–113.

Seidenberg, M. S., Waters, G., Sanders, M., & Langer, P. (1984). Pre- and post-lexical loci of contextual effects on word recognition. *Memory and Cognition, 12,* 315–328.

Shepard, L. (1980). An evaluation of the regression discrepancy method for identifying children with learning disabilities. *The Journal of Special Education, 14*(1), 79–91.

Siegel, L. S. (1989). IQ is irrelevant to the definition of learning disabilities. *Journal of Learning Disabilities,* 22(8), 469–478, 486.

Siegel, L. S. (1992). An evaluation of the discrepancy definition of dyslexia. *Journal of Learning Disabilities, 25*(10), 618–629.

Sinatra, G. M., & Royer, J. M. (1993). The development of cognitive component processing skills that support skilled reading. *Journal of Educational Psychology, 85,* 509–519.

Sinatra, G. M., & Royer, J. M. (1995, April). *Investigating a new method of combining response time and accuracy scores on measures of reading ability.* Paper presented at the annual meeting of the Society for the Scientific Study of Reading, San Francisco, CA.

Singley, M. K., & Anderson, J. R. (1989). *The transfer of cognitive skill.* Cambridge, MA: Harvard University Press.

Stanovich, K. E. (1986). Matthew effects in reading: Some consequences of individual differences in the acquisition of literacy. *Reading Research Quarterly, 21*(4), 360–406.

Stanovich, K. E. (1990). Concepts in developmental theories of reading skill: Cognitive resources, automaticity, and modularity. *Developmental Review, 10,* 72–100.

Stanovich, K. E. (1991a). Conceptual and empirical problems with discrepancy definitions of reading disability. *Learning Disability Quarterly, 14,* 269–280.

Stanovich, K. E. (1991b). Discrepancy definitions of reading disability: Has intelligence led us astray? *Reading Research Quarterly, 26*(1), 7–29.

Stanovich, K. E. (1993). A model for studies of reading disability. *Developmental Review, 13,* 225–245.

Stanovich, K. E., & Siegel, L. S. (1994). Phenotypic performance profile of children with reading disabilities: A regression-based test of the phonological-core variable-difference model. *Journal of Educational Psychology, 86*(1), 24–53.

van Dijk, T. A., & Kintsch, W. (1983). *Strategies for discourse comprehension.* New York: Academic Press.

Appendix

Sample Stimuli From the CAAS System

Task	Sample stimuli		
Simple	*** +++		
Letter	a, G, K, n		
Elementary-level word	you, goes, horse, banner		
Adult-level word	sprint, plight, kitten, canoe, baritone, pseudonym		
Elementary-level pseudoword	yob, poes, porse, danner		
Adult-level pseudoword	sprict, clight, fitken, yanob, larotine, psendinom		
Elementary-level category	YES: car/truck NO: bust/stool		
Adult-level category	YES: delegation/ballot NO: voter/gene		
Elementary-level semantics	The farmer *planted*/played the corn.		
Adult-level semantics	A district attorney's job is to *prosecute*/perpetrate the defendant.		
Phonological processing tasks	*Rhyme*	*Initial Sound*	*Final Sound*
Same sound/similar spelling	pain main	chain chair	size doze
Same sound/different spelling	shoe two	phase flush	trace lass
Different sound/similar spelling	food good	knit kite	cheese chess
Different sound and spelling	trip late	child open	niece splurge

Received October 30, 1996
Revision received March 17, 1997
Accepted March 31, 1997 ■