What lexical decision and naming tell us about reading

Leonard Katz · Larry Brancazio · Julia Irwin · Stephen Katz · James Magnuson · D. H. Whalen

Abstract The lexical decision (LD) and naming (NAM) tasks are ubiquitous paradigms that employ printed word identification. They are major tools for investigating how factors like morphology, semantic information, lexical neighborhood and others affect identification. Although use of the tasks is widespread, there has been little research into how performance in LD or NAM relates to reading ability, a deficiency that limits the translation of research with these tasks to the understanding of individual differences in reading. The present research was designed to provide a link from LD and NAM to the specific variables that characterize reading ability (e.g., decoding, sight word recognition, fluency, vocabulary, and comprehension) as well as to important reading-related abilities (phonological awareness and rapid naming). We studied 99 adults with a wide range of reading abilities. LD and NAM strongly predicted individual differences in word identification, less strongly predicted vocabulary size and did not predict comprehension. Fluency was predicted but with differences that depended on the way fluency was defined. Finally, although the tasks did not predict individual differences in rapid naming or phonological awareness, the failures nevertheless assisted in understanding the cognitive mechanisms behind these reading-related abilities. The results demonstrate that LD and NAM are important tools for the study of individual differences in reading.

Keywords Lexical decision · Naming · RT · Reading ability

L. Katz · J. Magnuson
Department of Psychology, University of Connecticut, Mansfield, CT, USA

L. Brancazio · J. Irwin
Department of Psychology, Southern Connecticut State University, New Haven, CT, USA

L. Katz (✉) · L. Brancazio · J. Irwin · S. Katz · J. Magnuson · D. H. Whalen
Haskins Laboratories, 300 George St., New Haven, CT 06511, USA
e-mail: leonard.katz@uconn.edu

Published online: 29 May 2011

Introduction

For decades, the most commonly used laboratory tasks for studying the cognitive processes involved in printed word identification have been lexical decision and naming (e.g., Jastrzembksi & Stanners, 1975; Rubenstein, Garfield, & Millikan, 1970; Zevin & Seidenberg, 2006). In the lexical decision (LD) task, the participant makes a speeded manual decision to a letter string on the computer screen: is it a word or not? In the naming task (NAM), the participant speaks aloud, as quickly as possible, a word that is printed on the screen. In both tasks, the measures of interest are the speed and accuracy of response. A typical experiment includes hundreds of such events or trials. Because responses are speeded, the process of identifying a word is automatic, not labored, and is thought to be similar in important ways to the word identification process in natural reading.

These tasks have been used to examine the characteristics that affect word identification, including word length, spelling regularity, neighborhood density, bigram frequency, word frequency, imageability, morphological transparency, orthographic depth and bilingualism (e.g., Frost & Katz, 1992). Also, LD and NAM have been employed frequently to assess models of printed word identification, lexical access, syntactic and semantic processing i.e., hypotheses and theories about the reading process (cf., Feldman & Andjelkovic, 1992). For example, LD and NAM tasks have been the main techniques used to study the extent of the involvement of phonology in printed word identification i.e., the importance of decoding-like processing (e.g., Lukatela & Turvey, 2000; Rastle & Coltheart, 2006). In recent years, the use of LD and NAM has been extended to neuroimaging studies of printed word processing to determine the locations of brain regions involved in normal reading and reading disability (for example, Fiebach, Friederici, Muller, & von Cramon, 2002; Frost et al., 2009; Graves, Desai, Humphreys, Seidenberg, & Binder, 2010; Pugh et al., 2006). They have also been used with electrophysiological investigations of printed word recognition in order to study the timing of the word identification process (e.g., Brown, Hagoort, & Chwilla, 2000).

In spite of the widespread use of LD and NAM tasks in studying fundamental aspects of the reading process, there has been little research to show how the tasks are performed by the reading disabled or more generally, how the tasks are related to specific measures of reading ability like word identification, fluency, comprehension, etc. In one of the few such studies, Marinus and deJong (2010) compared groups of Dutch fourth-grade dyslexic and typical readers in their sensitivity to various stimulus word characteristics (word length, frequency of appearance in reading material, and the number of "neighbor" words, that is, words whose spellings differ by only one letter). They found that dyslexics and beginning readers were more influenced by high frequency neighbors than were typical readers. However, other than defining a child as a good or poor reader, the authors did not collect more detailed information on component abilities such as comprehension, vocabulary, fluency, decoding, and sight word skill. Thus, although the Marinus and deJong study was substantially informative, we do not know, for example, if the dyslexics' weaker ability to inhibit an intrusive word neighbor (such as the word *miss* when *mist* was presented) was related to their weaker decoding functioning, smaller vocabulary, or inferior spelling ability. More generally—considering all levels of reading ability—we do not know which components of the reading process (decoding, sight word skill, fluency, vocabulary size, comprehension) are shared with the LD and NAM tasks and which are not. To this point, we have been unable to extend hundreds of laboratory studies on the structure of the mental lexicon and the word identification process to the understanding of reading ability differences.

The present research initiates an examination of the relationship between standardized reading assessments and the two laboratory tasks by asking several questions. First, is LD or NAM performance affected by a participant's decoding or sight word identification ability? Second, is reading fluency (the speed of automatic word recognition coupled with syntactic and semantic integration) correlated with performance speed in LD and NAM? We would expect a positive answer to both questions if the two tasks, in fact, are good models for the word identification process in natural reading (a common implicit assumption underlying the use of these paradigms in reading research).

A third issue is whether vocabulary size is correlated with performance in LD and/or NAM. Of course, a small vocabulary may lead to errors in the two tasks when the participant does not know the stimulus item. A more interesting question occurs when the words in the test are known to the participant (as in the present study): Does a larger vocabulary have an inhibitory effect on lexical retrieval because there is greater lexical competition from words that are similar to the stimulus item? Or is the opposite true: Larger vocabularies may facilitate retrieval (or correlate with mechanisms that afford efficient access)? Fourth, is response speed in LD and NAM a proxy for word identification speed as captured by reading fluency measures? Reading fluency in standardized reading tests assesses more than the automaticity of word identification; syntactic and semantic integration also play roles. If LD and NAM correlate with reading fluency measures, we will be reassured that these tasks, in fact, capture some degree of the same automaticity as natural reading. Further, it would also be of interest to know if LD and NAM are correlated with phonological awareness (PA), that is, with the ability to consciously manipulate the syllabic and phonemic content of speech. It is well known that PA is a predictor of reading development for alphabetic writing systems (Brady, Fowler, Stone, & Winbury, 1994; Cossu, Shankweiler, Liberman, Katz, & Tola, 1988; Torgesen, Wagner, Rashotte, Burgess, & Hecht, 1997) and, therefore, it is plausible that PA might relate to word identification in LD and NAM as well. Finally, rapid naming (RAN), a task that requires the rapid identification of colors, digits, letters, or pictures, is known to be a strong predictor of word identification in standardized reading tests (Georgiou, Parrila, Kirby, & Stephanson, 2008b; Scarborough, 1998; Wolf, Bowers, & Biddle, 2000), although why this is true is an issue still under discussion (Georgiou, 2008; Naples, Chang, Katz, & Grigorenko, 2009). The question for the present study is whether RAN is related to LD and NAM and, if so, why? Both LD and NAM require the rapid retrieval of words from lexical memory (Harm & Seidenberg, 2004; Rastle & Coltheart, 2006). If RAN also involves lexical retrieval, then we should see substantial correlations for RAN with LD and NAM. This would strengthen the hypothesis in which RAN is viewed as reflecting retrieval directly from the mental lexicon (Georgiou, Das, &

Hayward, 2008a). The opposing hypothesis proposes that RAN performance is based on an output from working memory and, therefore, is not directly affected by lexical factors (Arnell, Joanisse, Klein, Busseri, & Tannock, 2009). To address this question, as well as the questions related to the other components and correlates of reading ability that we have mentioned, this paper links what is known from laboratory experiments on word identification to what is known from standardized assessments of reading ability.

We enhanced the utility of the basic LD and NAM tasks by including in the test items both regularly and irregularly spelled words. It is known that irregular English words (e.g., *pint*) are spoken more slowly in NAM (Baron & Strawson, 1976; Glushko, 1979) and they tend to be recognized more slowly in LD (Seidenberg, Waters, Barnes, & Tanenhaus, 1984). For irregular words, pronunciation is slower because there is a common but incorrect way to pronounce the word (Castles et al., 2009). This regularity effect is also found with children (Waters, Seidenberg, & Bruck, 1984). The effect occurs when a transitory incorrect pronunciation of the irregular spelling pattern is mentally activated in the process of word identification. For example, the irregular *pint* shares an orthographic similarity but phonologic difference to the regular words *mint*, *lint*, *tint*, *hint*, and *dint*. If we find that irregular words like *pint* are slower than regular words (all differences other than regularity being equal), we can infer that misleading phonology has been activated in the process of identifying the item, presumably generated by decoding or analogy.

In one study that did directly compare LD and NAM on regular and irregularly spelled words, differences between responses to regular and irregular words were small or null in LD but greater in NAM (Katz et al., 2005). In addition to the behavioral observations, functional MRI (fMRI) measured activity in brain regions that are known to be important to the reading process. The fMRI results mirrored the behavioral differences between LD and NAM and they were consistent with the current understanding of the brain's major circuits for reading. Specifically, the NAM task activated the brain's dorsal circuit for decoding while the LD task activated the brain's fast-recognition occipital-temporal circuit that includes the visual word-form area (Paulesu et al., 2001; Fiebach, Friederici, Muller, & von Cramon, 2002). Thus, the fMRI data suggest that the NAM task promotes the generation of phonology for word identification more than the LD task. The LD task promotes processing at larger orthographic grain-sizes than the grapheme, e.g., at a multi-letter scale (Ziegler & Goswami, 2006).

As stated above, our main purpose was to relate LD and NAM to a detailed description of the components of reading ability. To achieve this, we administered a battery of reading and nonverbal IQ assessments (described below) as well as LD and NAM tasks. Our participants were drawn from programs for disabled students in universities and colleges in Connecticut and Rhode Island. Most of our participants had been diagnosed with a learning disability prior to the study. Consistent with this, many, but not all, tested below grade level in our own reading ability assessments. (First year college is equivalent to Grade 13.) We were interested in post-secondary school poor readers for two reasons. First, their reading difficulty is unlikely to be due solely to inadequate instruction or low intelligence (neither of which was of interest in this study). For even those primary school students who receive less than optimal instruction in decoding, those who are phonologically aware (i.e., the majority) typically teach themselves how to decode. Secondly, young adults would have the stamina and motivation to complete the hours of testing that our protocol required. Further, there is evidence that the deficits of adult poor readers are similar to those of younger readers; the study of one should produce insights into the other, with certain provisos. As Fowler and Scarborough (1993a, b) concluded in their review of adults with RD, "Their reading abilities appear to be hindered by weaknesses in the same components of the reading process that have been shown to pose the greatest challenges to children learning—and especially failing to learn—to read." Other studies on older poor readers agree in concluding that phonological deficit is the hallmark of RD in adults as well as in children (e.g., Bruck, 1992; Dietrich & Brady, 2001; Elbro, Nielsen, & Petersen, 1994).

Method

Participants

Potential participants were approached by email via the offices for disabled students that exist at most colleges and universities in Connecticut and were invited to volunteer for the study. Our intention was to form a sample that would be weighted toward below average readers but, nevertheless, represented a broad range of ability from very poor to very good. This was motivated by our decision to use correlational statistical techniques, which require substantial variability. Many of our participants were receiving accommodations (increased time on tests, etc.) for learning disabilities or specific reading disabilities; others had physical but not cognitive disabilities. The criteria for accommodations differed from school to school. Often accommodations were granted by a college disabilities office because the student had received them in primary and/or secondary school (although the bases for those earlier decisions were usually unknown). We expected and, in fact received, a wide range of reading scores but with the majority of students below average according to our own testing (described below). Means and medians were about 1/2 to 2/3 of a standard deviation below average for college students. The distributions of scores were unimodal and reasonably symmetric. Thus, as can be seen in Table 1, our sample was weighted toward poor readers.

Ninety-nine participants were recruited. They participated in exchange for payment and diagnostic feedback, receiving $14.00 for each hour of testing plus a bonus of $25 for completing the testing. In addition to the tests reported here, there were tests of speech production and perception and various control measurements; these are not reported here. The median age of subjects was 21.5 years. Nearly all of the subjects reported that they had difficulty reading although, as our standardized testing showed, their self-reports were not always reliable. All participants had adequate vision for reading as assessed by the MNRead Acuity Chart (1994).

Table 1 Correlations among NAM, LD, and simple RT

	NAM reg	NAM irreg	LD reg	LD irreg	LD Psw	Simple RT
NAM reg	1.00					
NAM irreg	0.88	1.00				
LD reg	0.61	0.43	1.00			
LD irreg	0.60	0.39	0.95	1.00		
LD Psw	0.70	0.57	0.83	0.82	1.00	
Simple RT	0.34	0.25	0.33	0.35	0.28	1.00
Mean ms	565.56	601.74	641.38	650.98	814.26	299.31
SD ms	120.02	138.8	165.62	178.85	281.85	62.52

Reg regular, *Irreg* irregular, *Psw* pseudoword

All correlations significant at $p < .05$ or less, $N_{\min} = 97$

Materials and procedure

We chose tests of reading ability that are widely used; therefore, their assets as well as their shortcomings are known to the community of reading researchers. These tests give measures of the components of reading ability such as decoding, sight word identification, reading fluency, comprehension, and more. Also, because we wanted to focus on speech factors relayed to reading, we choose instruments that test phonological awareness, phonological memory, and rapid naming. Additional measures were added as possible control variables of interest (measures of vocabulary, IQ, attention deficit, simple reaction time). For several factors (e.g., word identification, fluency, vocabulary, etc.), we had measurements from more than one test; we intended this to provide a validity and reliability check on our results.

Phonological abilities (phonological awareness, processing speed and memory) were assessed with Comprehensive Test of Phonological Processing (CTOPP; Wagner, Torgesen, & Rashotte, 1999), as well as reading and language assessment subtests within the Woodcock-Johnson III Diagnostic Reading Battery (WJ, Woodcock, Mather, & Schrank, 2004). The latter, along with the Test of Word Reading Efficiency, Form A (TOWRE; Torgesen, Wagner, & Rashotte, 1999) and the Gray Oral Reading Test-4, Form A (GORT, 2001) provided direct assessments of reading ability. For IQ assessment, all four subtests of the Wechsler Abbreviated Scale of Intelligence were administered (WASI; Wechsler, 1999). The Peabody Picture Vocabulary Test, Form A (PPVT; Dunn & Dunn, 2007) provided an index of vocabulary size along with subtests of the WJ and the WASI.

After the standardized tests, participants were given an LD task, a standard laboratory assessment of the speed and accuracy of printed word identification. On each trial, an asterisk appeared on the computer screen (the ready signal) replaced after 500 ms by a letter string. The participant pressed one of two computer keyboard keys as quickly as possible depending on whether their decision about the letter string was “word” or “not a word” (i.e., pseudoword). Maximum possible duration of the item was 3,000 ms; however, the presentation terminated as soon as a response was made. In between trials there was a blank screen for 1 s. There were three blocks of 40 trials each, 20 words and 20 pseudowords in quasi-random order. Ten of the words had regular spelling-to-sound correspondences (e.g., *bake*) and 10 were irregular (e.g., *bury*). We used a modest number of words because of the time constraints imposed by using a large number of assessments. Nevertheless, the words were chosen carefully to provide a representative set of moderately frequent regular and irregular words. The same pseudowords were repeated in blocks 2 and 3. Immediate repetition of any item was prohibited between blocks. Regularity was determined by conventional spelling-to-sound rules (cf., Venezky & Massaro, 1987; Ziegler, Stone, & Jacobs, 1997). All 20 words were moderately frequent in printed material (newspapers, books, magazines, etc.) according to Kucera-Francis (KF) norms (Kucera & Francis, 1967). The mean KF frequency for regular words was 7.6 per million and for irregular words, 15.7, for a task mean of 11.65. Twenty pseudowords were constructed that conformed to English orthotactic restrictions. The items are presented in Table 3 in the "Appendix". The task began with 16 practice items: 8 pseudowords and 8 moderately frequent words that did not appear in the main LD task. The participant's reaction time and accuracy were recorded after each trial.

Following the LD task, participants received the naming task. The sequence of events and timing were identical to LD with the exception that there were no pseudowords and the response that terminated a trial was the onset of vocalization. All 20 words in NAM were different from the words in the LD task. As with the LD stimuli, NAM stimuli were selected on the basis of their frequency and regularity. The mean KF frequency for regular words was 22.8 and for irregular words, 8.3, for a task mean of 15.5. There were 16 practice items, 8 regular words and 8 with irregular spelling; they were different from NAM task items and LD practice or task items. An analysis of variance on KF frequency comparing the 20 LD and 20 real NAM words (e.g., an analysis of task $\times$ regularity) showed no significant effects. In fact, effect sizes (squared partial etas) were very small: For task, eta-squared equaled .012; for regularity, .008; for their interaction, .094.

The subjects also participated in a simple (no choice) visual reaction time test in order to get a measure that could be used to refine choice response times by removing the effect of individual visual-motor differences. On each trial, participants were required to press a single key as soon as the stimulus, an asterisk, appeared. The asterisk always appeared in the center of the computer screen. Asterisks appeared at quasi-random intervals of 300, 500 and 700 ms. There were 15 trials at each of the intervals for a total of 45 trials.

Results

First we present separate internal analyses for lexical decision (LD), naming (NAM), the reading subtests, phonological awareness (PA) and rapid naming (RAN). Next we present analyses that relate LD and NAM to each other. Finally, we address the central issues of this paper, viz., the relationships between LD and

NAM, on the one hand, and the reading subtests with the reading-related variables PA and RAN, on the other.

Lexical decision

Accuracy

Percentage of correct responses was calculated for each participant in each combination of word type and block. Overall, accuracy in LD was higher for regular words than pseudowords. An analysis of variance was performed on accuracy as a function of word type (regular, irregular, pseudoword) and block (1, 2, 3). It produced a significant effect of word type × block, $F(4{,}98) = 7.54$, $p < .001$. Inspection of the interaction indicated that the percent of correct responses on regular and irregular word trials was nearly constant over blocks (for blocks 1–3, regular words: .94, .92, .92; irregular words: .88, .90, .90) while percent correct responses for pseudowords increased slightly, .86, .90, .90). Because LD accuracy data for regular and irregular words exhibited ceiling effects (small SDs), we did not use accuracy data in subsequent analyses.

Reaction time

Mean LD reaction time (RT) was calculated for each participant ($N = 99$) in each of the 9 combinations of word type and block. Each of the nine means for each participant was based on 10 data points (for regular or irregular words) or 20 data points (for pseudowords). RTs for incorrect responses were eliminated from the calculation. Figure 1 presents mean reaction time in milliseconds for regular, irregular, and pseudoword items over blocks. Inspection of the figure suggests that responses to pseudowords were slowest and that the greater decrease is from block 1 to block 2. The two real word conditions do not appear to differ. The data were subjected to an analysis of variance: word type (regular, irregular, pseudoword) × block (1, 2, 3). The suggestions were supported by the analysis: significant were word type, $F(2{,}202) = 68.24$, MSe = 41,296, $p < .001$; block, $F(2{,}202) = 16.00$, MSe = 15,136, $p < .001$; and word type × block, $F(4{,}404) = 4.37$, MSe = 8,395, $p < .003$. These results are similar to what was obtained by Katz et al. (2005) using different stimuli: Irregular words were recognized as quickly as regular. The slight apparent advantage for regular words versus irregular words in block 1 was not significant by a post hoc test in the present experiment (but was significant in Katz et al., 2005). We interpret this to indicate that decoding, which would have generated incompatible phonology for irregular words if it had been used, did not play a significant role in LD.

An items-analysis was also performed. For each of the 10 regular, 10 irregular and 20 pseudoword items, a mean RT for each block was calculated by averaging over participants. One advantage of the items-analysis is that it allowed us to assess the importance of three nuisance variables (covariates) that might also have affected response speeds. These were the (1) number of letters in an item, (2) word frequency of each item (pseudowords have frequency zero), and (3) orthographic

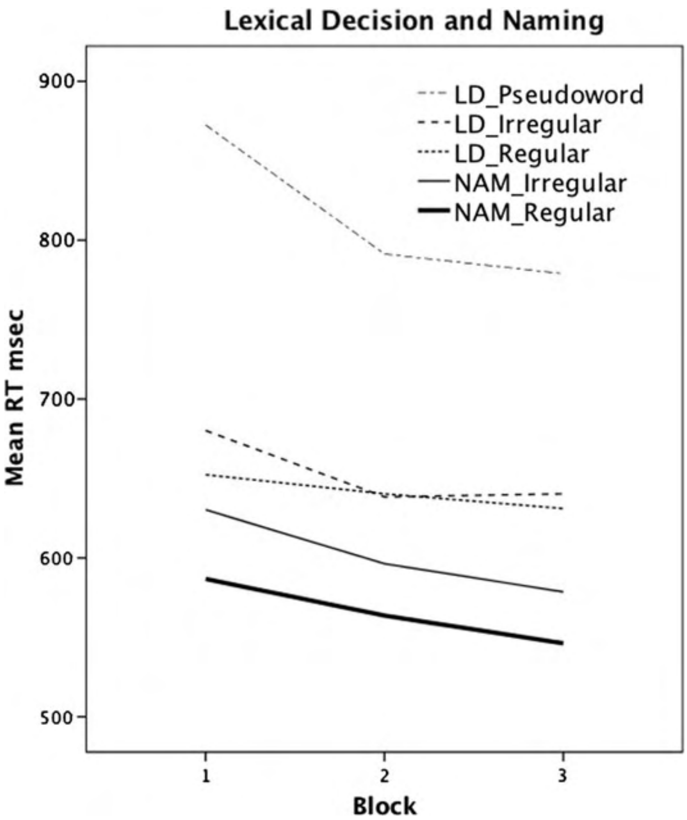

Fig. 1 Reaction time (*RT*) in milliseconds (*ms*) for lexical decision and naming tasks over three blocks of trials. Lexical decision conditions are in *broken lines*; naming in *solid lines*. *Reg* regular words, *Irr* irregular words, *Psw* pseudowords

neighborhood of each item (the number of real words that can be formed by replacing one letter at a time in the item). For word frequency we used the Kucera-Francis norms (Kucera & Francis, 1967). Items had either 4 or 5 letters, a small range for length; we did not expect to find an effect. Similarly, we did not expect to find the typical frequency effect (frequent words are recognized faster) because the range of frequency was deliberately restricted in our set of items. Orthographic neighborhood was indexed by the orthographic-N index from HAL (Balota et al., 2002). Words or pseudowords with larger orthographic-N ratings should have more lexical competition in the recognition process and, therefore, should be recognized more slowly but, again, our selection of items restricted the neighborhood range. However, although our stimuli had been selected to minimize these factors, some small differences between sets of items remained and the covariate item analysis was designed to assess the importance of this variation. An items analysis of covariance with the three nuisance variables showed that none of them was statistically significant and, in fact, their effects were negligible (all F values were less than unity). For the other sources of variance, the items analysis mirrored the previous subjects analysis, but with weaker significance levels.

Although regular and irregular words did not differ in LD, they did so in NAM (see below) so we maintained the regularity distinction in subsequent analyses in order to provide consistency across the two tasks. However, we reduced the data for each participant by averaging over his or her three blocks within each word type to produce a single mean for each word type.

Naming

Because the NAM task included no pseudowords and the words that were employed were frequent enough to be well known to participants (see Table 3 in the "Appendix"), errors were rare in block1, less than 1%. Because the same items were repeated in blocks 2 and 3, even fewer errors occurred in these latter blocks. Therefore, we made no statistical analysis of errors. Nevertheless, NAM responses for those few trials that were in error (mainly accidental early vocalizations and failures to respond) were eliminated from further analysis. Figure 1 presents NAM mean reaction times across blocks in milliseconds. The results contrast with the results for LD. It suggests that RT for irregular words was slower than regular words. Note that the initial disadvantage of irregular words continued in subsequent blocks.

These suggestions were confirmed by an analysis of variance on NAM RT for each participant in each combination of word type (regular, irregular) and block (1, 2, 3). Consistent with inspection of Fig. 1, word type was significant $F(1,98) = 29.89$, MSe = 6,502, $p < .001$. Thus, in contrast to LD, regular and irregular words in NAM did differ. Also significant was Block, $F(2,196) = 11.27$, MSe = 9,516, $p < .001$. Importantly, as Fig. 1 suggests, there was no significant interaction of word type × block. In fact, the mean square for the interaction was small, indicating a negligible effect size: $F(2,196) < 1$, $\mathrm{MS}_{\mathrm{wordtype} \times \mathrm{block}} = 1{,}995$, MSe = 5,256. It is clear that the difference between regular and irregular words was maintained over repetitions. Importantly, the result that word identification in LD is similar for irregular and regular words but dissimilar in NAM essentially replicates the result of Katz et al. (2005). In both studies, participants appear to generate phonology in the course of naming but not in lexical decision.

An item analysis was performed for NAM also. For each of the 10 regular and 10 irregular items, a mean RT for each block was calculated by averaging over participants. As with LD, the item analysis allowed us to assess the importance of three control covariates (the number of letters in an item, the word frequency of each item, and the orthographic neighborhood of each). An items analysis of covariance with the three control variables showed that for NAM, as for LD, the covariates were not statistically significant and, in fact, their effects were negligible (for all, $F(1,15) < 1$). When the analysis was run without the covariates, the results were similar to the subjects analysis: Significant effects for Block and Regularity but no significant interaction. Thus, the items analysis is consistent with the previous subjects analysis, albeit at weaker significance levels.

As was done for LD, RTs were aggregated over three blocks of trials to create one average score for each of the conditions. This produced more reliable data than maintaining separate block data and analyses carried out on the intact block data indicated that no important information was lost by averaging. Yap et al. (in press) found strong reliability in RT for both lexical decision and naming (although higher for naming) when measured either session-to-session or from even–odd item contrast.

LD and NAM

Table 1 presents correlations between the two NAM variables, the three LD variables, and simple RT. Inspection of Table 1 indicates high correlations within NAM and within LD and moderate correlations between the two sets. Only low nonsignificant correlations exist between simple RT and LD and NAM. Note, in Table 1, that $N_{\text{min}} = 97$ because some correlations were calculated on one or two fewer data points (missing data) than the maximum number of participants, $N = 99$.

Simple RT

It was clear from the nonsignificant correlations between simple RT and the reading subtests, PA, and RAN that simple RT accounted for no meaningful variance in any of them. And as demonstrated above, correlations between simple RT and all 5 LD and NAM variables were also not significant. Nevertheless, we followed a conservative strategy and continued to include simple RT as a control variable in all analyses so that the unique contributions of the LD and NAM variables will be unequivocally free from effects captured by simple RT.

Reading tests

Table 2 presents the means and standard deviations of selected Woodcock-Johnson Diagnostic Reading Battery III (WJ) subtests, Gray Oral Reading Test-4 (GORT), Test of Word Reading efficiency (TOWRE) and the Comprehensive Test of Phonological Processing (CTOPP). All data are standard scores, age adjusted, for their respective tests with $N = 99$. For each of the WJ subtests, the normative mean is 100 and the standard deviation (SD) is 15.

Inspection of Table 2 indicates that the WJ mean scores for the present sample were all below the normative mean except for the subtest sound blending. The mean score for all 10 subtests was 92.99, indicating that our sample generally performed lower than the population average for their college level (mean = 100, SD = 15). On the other tests, our sample means were also below average. Normative means and SDs are: TOWRE (100, 15), GORT (10, 3), CTOPP (10, 3), and WASI (10, 3). Inspection of Table 2 suggests that our sample's averages were well below the normative means except for vocabulary. Nevertheless, our sample displayed an adequate amount of individual difference variation. Correlations among the reading subtests, CTOPP, and WASI are also presented in Table 2. Note, however, that relationships among the reading and reading-related measures themselves are not the focus of the present paper.

Table 2 Correlations, means and SDs for standardized tests

	1	2	3	4	5	6	7	8	9	10	11	12	13
1. WJ word attack	1.00												
2. WJ letter word	0.57	1.00											
3. TOWRE sight word	0.27	0.19	1.00										
4. TOWRE phonemic decoding	0.62	0.45	0.51	1.00									
5. GORT accuracy	0.64	0.57	0.33	0.56	1.00								
6. WJ reading fluency	0.34	0.21	0.47	0.30	0.33	1.00							
7. GORT fluency	0.64	0.56	0.36	0.59	0.94	0.38	1.00						
8. WJ reading vocabulary	0.48	0.46	0.18	0.22	0.45	0.37	0.46	1.00					
9. WJ oral vocabulary	0.55	0.47	0.07	0.24	0.52	0.33	0.54	0.77	1.00				
10. WASI vocabulary	0.36	0.43	0.13	0.20	0.52	0.15	0.51	0.57	0.67	1.00			
11. PPVT	0.41	0.40	0.14	0.20	0.50	0.25	0.49	0.63	0.65	0.73	1.00		
12. WJ passage comprehension	0.43	0.34	0.25	0.24	0.41	0.36	0.40	0.64	0.55	0.34	0.49	1.00	
13. GORT comprehension	0.25	0.30	0.04	0.21	0.57	0.27	0.60	0.46	0.52	0.52	0.50	0.44	1.00
14. WJ sound awareness	0.36	0.15	0.09	0.11	0.08	0.15	0.08	0.33	0.40	0.28	0.09	0.20	0.04
15. WJ sound blending	0.36	0.21	0.23	0.24	0.23	0.19	0.23	0.33	0.35	0.29	0.23	0.30	0.16
16. CTOPP Blending Words	0.30	0.06	0.20	0.20	0.29	0.09	0.27	0.16	0.21	0.27	0.23	0.19	0.11
17. CTOPP blending nonwords	0.44	0.15	0.07	0.14	0.31	0.20	0.30	0.26	0.34	0.37	0.38	0.26	0.20
18. CTOPP elision	0.33	0.20	0.15	0.23	0.25	0.08	0.20	0.17	0.34	0.30	0.29	0.08	−0.05
19. CTOPP Phonological	0.59	0.47	0.16	0.41	0.51	0.14	0.51	0.34	0.51	0.45	0.39	0.29	0.27
20. CTOPP segmenting words	0.10	0.10	0.01	0.00	0.23	0.05	0.23	0.14	0.26	0.45	0.31	0.05	0.21
21. CTOPP segmenting nonwords	0.32	0.26	0.15	0.15	0.26	−0.02	0.25	0.17	0.25	0.31	0.34	0.19	0.11
22. CTOPP RAN objects	0.07	0.04	0.59	0.22	0.26	0.29	0.32	0.11	0.02	0.07	0.08	0.09	0.13
23. CTOPP RAN Letters	0.21	0.11	0.55	0.32	0.32	0.29	0.37	0.08	0.10	0.11	0.09	0.23	0.09
24. CTOPP RAN digits	0.20	0.03	0.54	0.26	0.25	0.27	0.28	0.06	0.04	0.05	0.04	0.20	0.07

Table 2 continued

	1	2	3	4	5	6	7	8	9	10	11	12	13
25. CTOPP RAN colors	0.04	0.10	0.44	0.21	0.23	0.32	0.26	0.13	0.03	0.01	0.18	0.22	0.19
Mean	90.98	93.49	82.29	85.87	9.55	83.33	7.93	95.09	96.98	54.70	102.90	87.68	7.74
SD	12.93	15.39	9.14	14.36	3.59	19.03	4.14	10.92	16.60	10.42	11.96	14.87	2.17

	14	15	16	17	18	19	20	21	22	23	24	25
15. WJ sound blending	0.49	1.00										
16. CTOPP blending words	0.30	0.53	1.00									
17. CTOPP blending nonwords	0.42	0.42	0.49	1.00								
18. CTOPP elision	0.37	0.21	0.32	0.47	1.00							
19. CTOPP phonological reversal	0.46	0.49	0.40	0.56	0.50	1.00						
20. CTOPP segmenting words	0.26	0.32	0.39	0.52	0.43	0.52	1.00					
21. CTOPP segmenting nonwords	0.30	0.46	0.44	0.57	0.36	0.47	0.67	1.00				
22. CTOPP RAN objects	0.09	0.25	0.28	0.10	0.13	0.17	0.14	0.14	1.00			
23. CTOPP RAN letters	0.00	0.05	−0.03	−0.04	0.09	0.20	0.07	0.10	0.51	1.00		
24. CTOPP RAN digits	0.00	0.11	0.01	0.04	0.14	0.21	0.15	0.17	0.48	0.82	1.00	
25. CTOPP RAN colors	−0.03	0.06	0.10	−0.03	0.05	0.09	0.08	0.05	0.56	0.46	0.44	1.00
Mean	93.20	102.42	10.07	10.23	8.53	8.65	9.86	8.88	8.48	8.26	9.11	9.03
SD	11.74	20.73	2.62	2.86	2.90	2.80	2.35	2.35	3.03	3.10	2.96	2.23

$N_{min} = 97$; $r > .198$; $p < .05$; $r > .258$; $p < .01$; $r > .320$, $p < .001$

Relating LD and NAM to reading tests

The next phase of analysis, our primary focus, examined the relationship of individual differences in LD and NAM performance to reading test measures: Decoding, sight word recognition, reading fluency, comprehension, and vocabulary. Because there are many variables relative to the number of participants, a hierarchical analysis strategy was implemented in order to reduce the number of statistical tests, thereby reducing the potential for type 1 errors. We began with canonical correlation, a technique in which a set of outcome variables can be regressed, simultaneously, on a set of predictor variables (Tabachnik & Fidel, 2001). Canonical correlation can be viewed as an extension of multivariate analysis of variance for the case in which both sides of the equation contain continuously scaled variables. We used this technique as a gatekeeper; subsequent analyses would explore only those relationships found to be statistically significant by the canonical correlation. In each canonical correlation, there were the same predictor variables: NAM regular RT, NAM irregular RT, LD regular RT, LD irregular RT, LD pseudoword RT, and as a control variable for basic response speed, simple RT. A canonical correlation was run for each of the following outcome factors: word reading, fluency, vocabulary, and comprehension. Two additional canonical correlations studied the reading-related factors, phonological awareness and rapid naming.

Word reading

Word Reading was represented by the following set of outcome subtest scores: WJ word attack, WJ letter word, TOWRE sight word, TOWRE phonemic decoding, and GORT accuracy. In some of these, performance had a time constraint. In the TOWRE, for example, participants are scored according to how many items are read correctly within 45 s. But for other subtests only accuracy is required (e.g., WJ word attack). As noted above, the predictor variables were: NAM regular RT, NAM irregular RT, LD regular RT, LD irregular RT, LD pseudoword RT, and simple RT.

Only the first root of the canonical correlation was significant, Wilk's Λ (30,346) = .4122, $p < .001$, with a quite substantial squared correlation of .462 between outcome and predictor variable sets. Because the overall canonical correlation was significant, the analysis advanced to conventional multiple regression. Each subtest was regressed separately on the set of predictors. The squared multiple correlations, R^2, for the individual regressions were: WJ word attack, .366, WJ letter word, .201, TOWRE sight word, .282, TOWRE phonemic decoding, .330, GORT accuracy, .277, all $p < .002$. WJ word attack and TOWRE phonemic decoding, both subtests that measure decoding skill directly, showed the strongest squared correlations.

Although the set of predictors related significantly to each subtest in the multiple regressions, there were only occasional significant individual predictors, indicating that particular predictor's unique contribution to subtest variance over and above the contribution mutually shared with the other predictors. When WJ letter word was regressed on the predictor set, no individual predictor was statistically significant.

For WJ word attack, only naming irregular RT was significant, $p < .001$. There was no significant single predictor for TOWRE sight word or for TOWRE phonemic decoding. For GORT accuracy, there were two significant predictors: Naming Irregular, $p < .005$, and LD pseudoword, $p = .031$.

To get further insight into the strength of the relationship of word reading to LD and NAM, we created a composite word reading score (the mean of all word reading subtest standard scores) for each participant. This composite was regressed on the three LD and two NAM predictors, plus simple RT; the result was an R^2 of .407, $p < .001$. The significant (unique) predictors were NAM irregular ($p < .002$) and LD pseudoword ($p = .025$). Two hierarchical (i.e., sequential) regressions were run to assess the unique contributions of LD and NAM. After entering simple RT into the regression, either the set of 3 LD variables was entered before the 2 NAM variables or vice versa. These showed that NAM accounted for substantial variance in the word reading composite over and above LD and simple RT: incremental $R^2 = .203$, $p < .001$. In contrast, the analogous incremental contribution of LD was small and not significant. Although both LD and NAM individually have significant relationships to word reading, it is clear that the stronger (and unique) relationship is for NAM. NAM, we have suggested, is largely a proxy for decoding skill. This conjecture was strengthened by hierarchical regressions on those subtests that emphasize either decoding or sight word skills (e.g., WJ word attack vs. WJ letter word, respectively). These showed that the superiority of NAM over LD was greater for the decoding subtests than for the sight word subtests.

Fluency

Fluency was represented by two outcome subtest scores: WJ reading fluency and GORT fluency. Both the first and second canonical roots were significant, respectively, Wilk's Λ $(12,176) = .5647$, $p < .001$, $R^2 = .310$, Wilk's Λ $(5,89) = .8186$, $R^2 = .181$, $p = .003$. Correlations between the roots (and their respective canonical variates) showed that root 1 corresponded mainly to the GORT and root 2 to the WJ. Therefore, WJ fluency and GORT fluency were not combined into a composite but were regressed separately on the set of six predictors. In spite of their differences, there is a moderate correlation between the two fluency measures ($r = .36$, $p < .001$, see Table 2). Nevertheless, there is a theoretical rationale for not combining them: the two reading tests differ in the way they measure fluency. In the WJ, both silent reading speed and the accuracy of the answer to a comprehension question at the end of each sentence contribute to a participant's fluency score. In contrast, for GORT fluency the score is a combination of oral reading rate and spoken word accuracy; comprehension is assessed only in a separate measurement at the end of each paragraph and does not contribute to the fluency score.

The squared multiple correlations were moderate: for WJ fluency, $R^2 = .236$, and for GORT fluency, $R^2 = .297$, both $p < .001$. The only significant unique predictors of WJ fluency were LD Regular ($p < .002$) and LD Irregular ($p = .015$). For GORT fluency, there were two predictors that showed unique (albeit weaker) effects, NAM Irregular ($p = .043$) and LD pseudoword ($p = .025$).

In order to assess the relative strengths of LD and NAM as predictors of WJ fluency and GORT fluency, we ran hierarchical regressions in which the set of LD variables was entered after the set of NAM variables or vice versa. Every hierarchical regression entered simple RT first and then either the 2 NAM variables or the 3 LD variables. For WJ fluency, the incremental R^2 for LD over NAM was .13, $p < .002$. The incremental R^2 for NAM over LD was .01, not significant. But when the outcome was GORT fluency, LD failed to predict significant variance above NAM, $R^2 = .04$. However NAM, when it was entered last, had an incremental $R^2 = .11$, $p < .001$. As was the case for the preceding regressions for fluency, these results indicate that LD is a better indicator of fluency in the Woodcock-Johnson and that NAM is a better indicator of fluency for the Gray Oral Reading Test. At the least, it seems prudent to suggest that WJ fluency and GORT fluency are not assessing the same skills.

Vocabulary

Vocabulary was represented by the following subtest outcome variables: WJ reading vocabulary, WJ oral vocabulary, WASI vocabulary, and PPVT. The predictor variables were the same as above. Only the first canonical root was significant, Wilk's Λ (24,297) $= .6177$, $p = .011$, with a moderate squared correlation of .237 between outcome and predictor variable sets.

Each vocabulary outcome variable was regressed, individually, on the set of six predictors. The squared multiple correlations for the individual subtests were: WJ reading vocabulary, $R^2 = .135$, $p = .043$, WJ oral vocabulary, $R^2 = .160$, $p = .015$, WASI vocabulary, $R^2 = .217$, $p < .001$, and PPVT, $R^2 = .216$, $p < .001$. LD Irregular and LD pseudoword RT were significant unique predictors for WJ oral vocabulary ($p = .01$, $p = .02$, respectively). For WASI vocabulary, only simple RT was significant ($p = .01$). Finally, for PPVT, only LD irregular was significant ($p = .01$). These results suggest that LD is more strongly related to vocabulary (both reading vocabulary and oral vocabulary) than is NAM, although neither is a strong predictor.

A composite score for each participant was created representing the mean of all subtest standard scores (rescaled for comparability). When it was regressed on the three LD and two NAM predictors plus simple RT, R^2 equalled .209. In hierarchical regressions, in order to assess the relative contributions of LD and NAM to vocabulary, the incremental R^2 for LD over NAM was .09, $p < .02$. However, NAM failed to provide any significant incremental increase in R^2 beyond LD: $R^2 = .04$, ns. Again, the results indicate that LD is the task that is more sensitive to vocabulary size.

Comprehension

Reading comprehension was represented by WJ passage comprehension and GORT comprehension. The predictor variables were the same as above. The canonical correlation was not significant, Wilk's Λ (12,178) $= .8888$, $p = .548$. LD and NAM do not predict reading comprehension.

Phonological awareness

Phonological awareness was represented by WJ sound awareness, WJ sound blending, and the following CTOPP subtests: blending words, blending nonwords, elision, phoneme reversal, segmenting words, and segmenting nonwords. The predictor variables were the same as previously used. The canonical correlation was not significant, Wilk's Λ (48,412) = .4968, $p = .086$. Neither LD nor NAM appeared to be related to phonological awareness. This null result does not appear to be a function of low variability in the subtests (see Table 2 for descriptive statistics).

A null result was unexpected because PA is known to correlate strongly with word reading and, as we have already seen, the LD and NAM tasks also correlate with word reading. A simple explanation is that the two tasks and PA account for different aspects of word reading, that is, that they do not overlap. In order to assess this explanation, we ran hierarchical regressions in which the word reading composite described above was regressed on both the task variables (LD, NAM), and PA. The combination produced an R^2 of .472 over and above simple RT. Whether the task variables were entered before PA or the order was reversed, both final regression stages accounted for significant incremental variance over the other indicating that some independence exists between the laboratory tasks and PA in accounting for word identification. The final stage contribution of the combined LD and NAM variables to word reading variance over and above PA was large ($R^2 = .318$, $p < .001$), compared to the increment attributed to PA when it was entered last ($R^2 = .033$, $p < .008$). It is clear then, in the present data, that (1) although the tasks do not correlate with PA, and (2) both the tasks and PA correlate with word reading, (3) the tasks account for considerable variance in word reading that is independent of PA.

Rapid naming

The rapid naming outcome variables were the four CTOPP subtests for naming objects, letters, digits, and colors. The canonical correlation between subtests and task predictors (and including simple RT) was not significant, Wilk's Λ (24,304) = .6849, $p = .081$. This null result is surprising because, on an intuitive task analysis of RAN, LD, and NAM, all three can be viewed as involving the fast retrieval of words from lexical memory. We ran hierarchical regressions similar to the analyses for PA, in which the word reading composite was regressed on LD and NAM, in one stage, and a RAN composite formed by the mean of the four RAN subtests in a second stage. When RAN preceded LD and NAM, the latter stage accounted for an R^2 change of .307 ($p < .001$). However, when RAN constituted the final stage, the incremental R^2 change was only a nonsignificant .021. Thus, it appears not only that RAN does not relate to LD and NAM but RAN does not account for any substantial variance in word reading over and above the contribution of LD and NAM. We discuss possible interpretations of this null result below.

Discussion

The lexical decision (LD) and naming (NAM) tasks are often used to test models of printed word identification, lexical access, syntactic and semantic processing—hypotheses and theories of the reading process. The present research was designed to provide a link between these ubiquitous laboratory tasks and variables that characterize reading ability (e.g., decoding, sight word identification, fluency, vocabulary size, and comprehension) as well as two reading-related abilities (phonological awareness and rapid naming). We found several strong relationships between the tasks and reading ability. Just as interesting, there were some informative failures to find relationships that had seemed, a priori, to be likely. In all, our results support the often implicit assumption that LD and NAM provide useful proxies for reading ability and the results refine that assumption with regard to specific skills.

In particular, the results showed very robust relationships between participants' word identification test scores and lexical decision and naming. When single word-reading accuracy was assessed by canonical correlation, LD and NAM together accounted for a substantial portion (46%) of the variance in the set of word identification subtests WJ word attack, WJ letter word, TOWRE sight word, TOWRE phonemic decoding, and GORT accuracy. Further analyses indicated that NAM accounted for unique variance in word identification over and above LD (but not vice versa) and the strongest relationship was between naming and decoding ability (i.e., word attack and phonemic decoding). Thus, naming is a good index of decoding ability.

Comparisons between LD and NAM with regard to spelling regularity showed that irregular words were named more slowly than regular words in NAM but this result was not found in LD. This result replicated earlier findings (Katz et al., 2005) and it suggests, again, that the naming process depended more on decoding than did LD.

We found that poor readers were slower to correctly name irregular words. In five separate regressions, each of the word reading subtests was regressed on the six predictors (RT for the three LD word types, regular, irregular, and pseudoword; RT for two NAM variables, regular and irregular, and simple RT). None of the predictors accounted for any unique variance in any of the reading subtests except for the speed of naming irregular words and, to a lesser extent, lexical decision speed to pseudowords. Irregular word naming stresses processing more than the other word types because there is a conflicting but incorrect regular way to pronounce the word (Castles et al., 2009). The poorer readers' lower skill level could be the result of weaker decoding skills or less reading experience; both factors are likely to be involved. A similar explanation can account for the unique correlation between lexical decision speed and pseudowords. Inexperienced readers may be more likely to process peudowords in LD by means of decoding; their slow speed in doing so will correlate with their poorer scores in the word reading subtests.

Simple RT (visual-motor response time) had a negligible relation to word reading and to any of the other reading or reading related measures we studied. The absence of a simple RT effect on measures of reading ability has been found with children as well as, in the present study, with adults. It is consistent with other studies that have found that poor reading is not the consequence of slower visual-motor processing (cf., Katz & Wicklund, 1971; Naples, Katz, & Grigorenko, submitted). Stringer and Stanovich (2000) did find some individual difference effects for response speed (although still small) but only when the task was a two-choice decision, not simple RT.

With regard to reading fluency, LD and NAM together predicted WJ fluency and GORT fluency moderately well with $R^2 = .236$ for WJ and $R^2 = .297$ for GORT. Although both tasks were correlated with both fluency subtests, hierarchical regressions revealed that the LD task was clearly a better predictor of the Woodcock-Johnson and NAM was the better predictor for the GORT. This may be explained, perhaps, by the different ways that fluency is assessed by the WJ and the GORT. First, comprehension is involved in the WJ's fluency measure; participants must answer a comprehension question at the end of each of the sentences they read thus including the time to answer the question in the time taken to read the sentence. For the GORT, comprehension questions are answered only after the fluency measure has already been determined by the accuracy and rate of reading aloud. Another difference is that participants perform the WJ fluency test while reading silently but perform the GORT by reading aloud. The lexical decision task, like the WJ fluency test, requires silent reading along with a manual response. On the other hand, the naming task, like the GORT, requires a vocal response. Katz et al. (2005) suggest that the requirement to speak aloud (as in the naming task) stimulates parts of the brain's reading circuit (the dorsal circuit) that specialize in decoding. The alternative circuit, which contains the visual word form area is thought to process larger "grain size" units of print more efficiently (larger than individual letters, Ziegler & Goswami, 2006).

Regressing vocabulary on LD and NAM gave clear evidence that it is the LD task that has the stronger relationship to vocabulary size, although neither task had more than a modest correlation. We regard vocabulary size (both oral and reading vocabulary) as a proxy for reading experience; experience with print is likely to increase vocabulary size (Yap et al., in press). The greater a participant's vocabulary, the faster were their responses in LD and NAM. One could have argued a priori that the opposite might obtain: Large vocabularies might have created interference or competition in the mental lexicon when a participant attempts to recognize or name the target word. But the opposite is the case; readers with larger vocabularies also have faster retrieval processing, a result also found by Yap et al.

With regard to reading comprehension, it is clear that neither LD nor NAM is informative. In one sense, this is not surprising because neither task requires reading comprehension. On the other hand, because comprehension is correlated with variables such as word accuracy, fluency, and vocabulary—all of which are, in turn, correlated with LD and NAM—it was possible that mediated correlations might exist. Furthermore, it is undeniable that word identification is a necessary prerequisite for comprehension; lexical knowledge is central (Perfetti, 2007). That we do not find such a connection between the tasks and comprehension simplifies our picture of the skills captured by LD and NAM. Again, we must qualify our finding by the fact that our participants were adults; a different picture might emerge with younger readers. The reaction time measure central to LD and NAM is suitable only when participant accuracy is nearly perfect. If participants do not have the words that are used in LD and NAM in their lexicons prior to the testing session, use of the tasks is problematic.

Neither phonological awareness (PA) nor rapid naming (RAN) had significant canonical correlations with LD or NAM. To consider PA first, it is well known that PA is a major predictor of word reading in children (Scarborough, 1998). This was true as well for the present study's adult participants; some measures of reading correlated significantly with some measures of PA. For example, Table 2 shows that WJ word attack correlated with the phonological awareness-related subtests of the CTOPP, blending words ($r = .265$), blending nonwords ($r = .445$), elision ($r = .244$), and phoneme reversal ($r = .531$); all $p < .015$). That we failed to find a significant canonical correlation between the set of LD and NAM variables and the set of aforementioned CTOPP variables (as well as all other CTOPP variables) suggests that performance in LD and NAM, for adults at least, is not related to performance on PA tasks. Hierarchical regressions supported this suggestion. Further, although both PA and the two laboratory tasks accounted for significant proportions of word reading variance, we found that the contribution of the tasks to word reading, although substantial, was largely independent from the contribution of PA to word reading. Thus, we conclude that LD and NAM do not share the kind of processing captured by the CTOPP's PA subtests. Again, we point out that although this may be true for our participants who are adults, it may be different for children. Scarborough (2010) has pointed out that any specific marker of poor reading may be prominent at one age but not so at another.

Although we are able to rationalize why PA is not related to the two tasks, the finding that RAN is also uncorrelated is perhaps more counterintuitive. LD, NAM, and RAN could all be viewed as requiring fast access to the mental lexicon. In all three, words have to be retrieved in response to a print stimulus (as in LD and NAM) or pictures and symbols (as in RAN). The literature on LD supports the assumption that, in order to execute a lexical decision, the mental lexicon must be activated. The lexical decision task is frequently employed to study how a word is retrieved from the lexicon (e.g., Rueckl & Aicher, 2008). For NAM, lexical access is not required for words that have a regular spelling and therefore can be pronounced according to rule, but lexical access is clearly necessary to pronounce irregular words correctly. In the present study's NAM task, half the items were irregularly spelled words; therefore, it would have been prohibitive for participants to avoid lexical access. Their very high accuracy rate argues that they did not avoid it. The literature contains many data and modeling papers that are based on the assumption that lexical access typically occurs in NAM as well as LD (e.g., Harm & Seidenberg, 2004; Coltheart, Rastle, Perry, & Langdon, 2001; Plaut, McClelland, Seidenberg, & Patterson, 1996). So, it seems to us that there is every reason to believe that the LD task involves lexical access. Therefore, we propose that the proximal process that accounts for response speed in RAN does not involve retrieval from the mental lexicon itself but rather retrieval from a working memory. In this short term buffer, RAN items are held ready for rapid output. We will not speculate on the characteristics of this short term store but see Arnell et al. (2009) and Katz and Shankweiler (1985), who propose a similar notion and Georgiou, Das, and Hayward (2008a), who discuss the role of working memory in both RAN and PA.

In summary, this paper demonstrates that LD and NAM provide substantial information correlated with reading ability and, in particular, are good paradigms for studying individual differences in word identification (both sight word and decoding processes) and fluency. Future work will focus first on the design of efficient and reliable versions of LD and NAM that can be adapted for use across the age range. Then, because LD and NAM are proven basic instruments for investigating the cognitive processes of word identification and lexical access, they can be employed in experiments that study the finer structure of individual differences in reading ability.

Acknowledgments This work was supported by NIH grant HD-01994 to Haskins Laboratories.

Appendix

See Table 3.

Table 3 LD and NAM items

LD reg	*LD pseudo*
Dock	Binch
Flag	Brone
Float	Chire
Goat	Gwill
Greed	Leard
Heel	Phash
Hike	Plove
Junk	Preed
Rust	Sheik
Stack	Slace
LD irreg	Slock
Choir	Stelk
Cough	Stroat
Flood	Tayes
Glove	Theel
Gross	Tound
Isle	Tross
Palm	Wark
Shoe	Yade
Swan	Yooly
Worm	
NAM reg	*NAM irreg*
Bake	Bury

Table 3 continued

Brake	Comb
Chew	Deaf
Clay	Hymn
Grill	Pear
Hen	Pint
Limp	Plow
Pest	Sew
Slick	Tomb
Wagon	Wasp

Reg regular, *Irreg* irregular

References

Arnell, K. M., Joanisse, M. F., Klein, R. M., Busseri, M. A., & Tannock, R. (2009). Decomposing the relation between rapid automatized naming (RAN) and reading ability. *Canadian Journal of Experimental Psychology, 63*, 173–184.

Balota, D. A., Cortese, M. J., Hutchison, K. A., Neely, J. H., Nelson, D., Simpson, G. B., et al. (2002). The English lexicon project: A web-based repository of descriptive and behavioral measures for 40,481 English words and nonwords. Retrieved from http://elexicon.wustl.edu.

Baron, J., & Strawson, C. (1976). Use of orthographic and word-specific knowledge in reading words aloud. *Journal of Experimental Psychology: Human Perception and Performance, 2*, 386–393.

Brady, S., Fowler, A., Stone, B., & Winbury, N. (1994). Training phonological awareness: A study with inner-city kindergarten children. *Annals of Dyslexia, 44*, 26–59.

Brown, C. M., Hagoort, P., & Chwilla, D. J. (2000). An event-related brain potential analysis of visual word priming effects. *Brain and Language, 72*, 158–190.

Bruck, M. (1992). Persistence of dyslexics' phonological awareness deficits. *Developmental Psychology, 28*(5), 874–886.

Castles, A., Coltheart, M., Larsen, L., Jones, P., Saunders, S., & Mcarthur, G. (2009). Assessing the basic components of reading: A revision of the Castles and Coltheart test with new norms. *Australian Journal of Learning Difficulties, 14*, 67–88.

Coltheart, M., Rastle, K., Perry, C., & Langdon, R. (2001). DRC: A dual route cascaded model of visual word recognition and reading aloud. *Psychological Review, 108*, 204–256.

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

Dietrich, J. A., & Brady, S. A. (2001). Phonological representations of adult poor readers: an investigation of specificity and stability. *Applied Psycholinguistics, 22*, 383–418.

Dunn, L. M., & Dunn, D. M. (2007). *Peabody picture vocabulary test* (4th ed.). Minneapolis, MN: Pearson.

Elbro, C., Nielsen, I., & Peterson, D. K. (1994). Dyslexia in adults: Evidence for deficits in non-word reading and in the phonological representation of lexical items. *Annals of Dyslexia, 44*, 203–226.

Feldman, L. B., & Andjelkovic, D. (1992). Morphological analysis in word recognition. In R. Frost & L. Katz (Eds.), *Orthography, phonology, morphology, and meaning* (pp. 343–361). Amsterdam: Elsevier.

Fiebach, C. J., Friederici, A. D., Muller, K., & von Cramon, D. Y. (2002). fMRI evidence for dual routes to the mental lexicon in visual word recognition. *Journal of Cognitive Neuroscience, 14*, 11–23.

Fowler, A. E., & Scarborough, H. S. (1993). *Should reading disabled adults be distinguished from other adults seeking literacy instruction? A review of theory and research*. Technical report no. 93-7, National Center for Adult Literacy, University of Pennsylvania.

Fowler, A. E., & Scarborough, H. S. (1993). *Should reading-disabled adults be distinguished from other adults seeking literacy instruction? A review of theory and research*. Report no: NCAL-TR-93-7. National Center on Adult Literacy.

Frost, R., & Katz, L. (Eds.). (1992). *Orthography, phonology, morphology, and meaning*. Amsterdam: Elsevier.

Frost, S. J., Landi, N., Mencl, W. E., Sandak, R., Fulbright, R. K., Tejada, E.T., et al. (2009). Phonological awareness predicts activation patterns for print and speech. *Annals of Dyslexia*, *59*, 78–97

Georgiou, G. (2008). Why is rapid naming speed related to reading? Examining different theoretical accounts. Ph.D. dissertation, University of Alberta.

Georgiou, G., Das, J. P., & Hayward, D. V. (2008a). Comparing the contribution of two tests of working memory to reading in relation to phonological awareness and rapid naming speed. *Journal of Research in Reading, 31*, 302–318.

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

Glushko, R. J. (1979). The organization and activation of orthographic knowledge in reading aloud. *Journal of Experimental Psychology: Human Perception and Performance, 5*, 674–691.

Graves, W. W., Desai, R., Humphreys, C., Seidenberg, M. S., & Binder, J. R. (2010). Neural systems for reading aloud: A multiparametric approach. *Cerebral Cortex, 20*, 1799–1815.

Gray Oral Reading Test. Revision 4. (2001). Austin, TX: Pro-ed.

Harm, M. W., & Seidenberg, M. S. (2004). Computing the meanings of words in reading: Cooperative division of labor between visual and phonological processes. *Psychological Review, 111*, 662–720.

Jastrzembski, J. E., & Stanners, R. F. (1975). Multiple word meanings and lexical search speed. *Journal of Verbal Learning and Verbal Behavior, 14*, 534–537.

Katz, L., Lee, C. H., Tabor, W., Frost, S. J., Mencl, W. E., Sandak, R., et al. (2005). Behavioral and neurobiological effects of printed word repetition in lexical decision and naming. *Neuropsychologia, 43*, 2068–2083.

Katz, R., & Shankweiler, D. (1985). Repetitive naming and the detection of word retrieval deficits in the beginning reader. *Cortex, 21*, 617–625.

Katz, L., & Wicklund, D. (1971). Simple reaction time for good and poor readers in grades two and six. *Perceptual and Motor Skills, 32*, 270.

Kucera, H., & Francis, W. N. (1967). *Computational analysis of present-day American English*. Providence, RI: Brown University Press.

Lukatela, G., & Turvey, M. T. (2000). An evaluation of the two-cycles model of phonology assembly. *Journal of Memory and Language, 42*, 183–207.

Marinus, E., & deJong, P. F. (2010). Size does not matter, frequency does: Sensitivity to orthographic neighbors in normal and dyslexic readers. *Journal of Experimental Child Psychology, 106*, 129–144.

MNRead Acuity Chart. (1994). *Regents of the University of Minnesota*.

Naples, A. J., Chang, J. T., Katz, L., & Grigorenko, E. (2009). Same or different? Insights into the etiology of phonological awareness and rapid naming. *Biological Psychology, 80*, 226–239.

Naples, A. J., Katz, L., & Grigorenko, E. L. (submitted). *A diffusion model analysis of reaction time and reading skill*.

Paulesu, E., Demonet, J.-F., Fazio, F., McCrory, E., Chanoine, V., Brunswick, N., et al. (2001). Dyslexia: Cultural diversity and biological unity. *Science, 291*, 2165–2167.

Perfetti, C. A. (2007). Reading ability: Lexical quality to comprehension. *Scientific Studies of Reading, 11*, 357–383.

Plaut, D. C., McClelland, J. L., Seidenberg, M. S., & Patterson, K. E. (1996). Understanding normal and impaired word reading: Computational principles in quasi-regular domains. *Psychological Review, 103*, 56–115.

Pugh, K. R., Sandak, R., Frost, S. J., Moore, D., Rueckl, J. G., & Mencl, W. E. (2006). Neurobiological studies of skilled and impaired reading: A work in progress. In G. D. Rosen (Ed.), *The dyslexic brain: New pathways in neuroscience discovery* (pp. 21–47). Mahwah, NJ: Erlbaum.

Rastle, K., & Coltheart, M. (2006). Is there serial processing in the reading system; and are there local representations? In S. Andrews (Ed.), *From inmarks to ideas: Current issues in lexical processing* (pp. 3–24). Hove, UK: Psychology Press.

Rubenstein, H., Garfield, L., & Millikan, J. (1970). Homographic entries in the internal lexicon. *Journal of Verbal Learning and Verbal Behavior, 9*, 487–494.

Rueckl, J. G., & Aicher, K. A. (2008). Are CORNER and BROTHER morphologically complex? Not in the long term. *Language and Cognitive Processes, 23*, 972–1001.

Scarborough, H. S. (1998). Predicting the future achievement of second graders with reading disabilities: Contributions of phonemic awareness, verbal memory, rapid naming, and IQ. *Annals of Dyslexia, 48*, 115–136.

Scarborough, H. S. (2010). *When matters! Keynote address to the meeting of the Society for the Scientific Study of Reading*. Berlin

Seidenberg, M. S., Waters, G., Barnes, M., & Tanenhaus, M. K. (1984). When does irregular spelling or pronunciation influence word recognition? *Journal of Verbal Learning and Verbal Behavior, 23*, 383–404.

Stringer, R., & Stanovich, K. E. (2000). The connection between reaction time and variation in reading ability: Unraveling covariance relationships with cognitive ability and phonological sensitivity. *Scientific Studies of Reading, 4*, 41–53.

Tabachnik, B. G., & Fidel, L. S. (2001). *Using multivariate statistics* (4th ed.). Boston, MA: Allyn and Bacon.

Torgesen, J. K., Wagner, R. K., & Rashotte, C. A. (1999). *Test of word reading efficiency*. Austin, TX: Pro-ed.

Torgesen, J. K., Wagner, R. K., Rashotte, C. A., Burgess, S., & Hecht, S. (1997). Contributions of phonological awareness and rapid automatic naming ability to the growth of word-reading skills in second-to fifth-grade children. *Scientific Studies of Reading, 1*, 161–185.

Venezky, R. L., & Massaro, D. W. (1987). Orthographic structure and spelling-sound regularity in reading English words. In A. Allport, D. MacKay, W. Prinz, & E. Scheerer (Eds.), *Language perception and production: Shared mechanisms in listening, speaking, reading, and writing* (pp. 158–179). London: Academic Press.

Wagner, R. K., Torgesen, R. K., & Rashotte, C. (1999). *Comprehensive test of phonological processing*. Austin, TX: Pro-ed.

Waters, G. S., Seidenberg, M. S., & Bruck, M. (1984). Children's and adults' use of spelling-sound information in three reading tasks. *Memory and Cognition, 12*, 293–305.

Wechsler, D. (1999). *Manual for the Wechsler Abbreviated Scale of Intelligence*. San Antonio, TX: The Psychological Corporation.

Wolf, M., Bowers, P. G., & Biddle, K. (2000). Naming speed processes, timing, and reading: a conceptual review. *Journal of Learning Disabilities*, *33*, 387–407

Woodcock, R. W., Mather, N., & Schrank, F. A. (2004). *Woodcock-Johnson III diagnostic reading battery*. Itasca, IL: Riverside Publishing.

Yap, M. J., Balota, D. A., Sibley, D. E., & Ratcliff, R. (In Press). Individual differences in visual word recognition: Insights from the ELP. *Journal of Experimental Psychology: Human Perception and Performance*.

Zeigler, J. C., & Goswami, U. (2006). Becoming literate in different languages: Similar problems, different solutions. *Developmental Science, 9*, 429–453.

Zevin, J. D., & Seidenberg, M. S. (2006). Simulating consistency effects and individual differences in nonword NAM: A comparison of current models. *Journal of Memory and Language, 54*, 145–160.

Ziegler, J. C., Stone, G. O., & Jacobs, A. M. (1997). What's the pronunciation for _OUGH and the spelling for /u/? A database for computing feedforward and feedback inconsistency in English. *Behavior Research Methods, Instruments, & Computers, 29*, 600–618.