American Psychological Association

Journal of Educational Psychology

0022-0663/19/$12.00

2019, Vol. 111, No. 2, 314–330
http://dx.doi.org/10.1037/edu0000279

Automaticity of Word Recognition Is a Unique Predictor of Reading Fluency in Middle-School Students

Tanja C. Roembke, Eliot Hazeltine, Deborah K. Reed, and Bob McMurray
University of Iowa

Automaticity in word recognition has been hypothesized to be important in reading development (LaBerge & Samuels, 1974; Perfetti, 1985). However, when predicting educational outcomes, it is difficult to isolate the influence of automatic word recognition from factors such as processing speed or knowledge of grapheme-phoneme correspondences. Cognitive models suggest automaticity could be achieved in different components of word recognition (e.g., by memorizing familiar words or by tuning the mappings between orthography, phonology or semantics). However, the contributions of each path have not been assessed. This study developed a new measure of automaticity to overcome these limitations and relates automaticity to standard outcomes. Subjects were 58 middle-school students (mean age = 13.2 years ± 8 months) with average to below-average reading comprehension. To assess automaticity with an accuracy-based measure, backward masking was used: On half the trials, items were presented for 90 ms and replaced by a nonlinguistic mask; on the other half it was presented unmasked to assess children's knowledge of the word. This was instantiated in 3 experimental tasks developed to maximize reliance on different reading mappings. Automaticity, particularly in a task stressing meaning, predicted reading fluency over and above knowledge of the relevant grapheme-phoneme mappings. Automaticity in tasks involving nonwords also predicted fluency, suggesting the possibility of automaticity in orthography to phonology mappings. Variation in automatic word recognition did not predict reading comprehension or decoding. This link between automatic written word recognition and fluency has important implications for how automaticity may be targeted to improve reading outcomes.

Educational Impact and Implications Statement

This study asked whether the degree to which children recognize written words automatically predicts reading fluency, decoding, and comprehension in low-performing middle-school students. Prior work has not developed measures of automaticity that are independent of factors such as knowledge of the words and letters or general differences in speed of processing. The authors assessed automaticity by briefly presenting a written word and then covering it with a visual mask to force children to process the word rapidly; this is called backward masking. Students' levels of automaticity did not predict their ability to decode words or comprehend what they read; however, performance under masked conditions predicted reading fluency over and above their knowledge of the words and letters. These results suggest that developing automaticity in word recognition may be an important factor in middle-school students' struggles and, thus, a prime target for intervention. Backward masking may offer an innovative way to assess automaticity to identify these students.

Keywords: reading, automaticity in word recognition, fluency

Supplemental materials: http://dx.doi.org/10.1037/edu0000279.supp

This article was published Online First May 3, 2018.

Tanja C. Roembke and Eliot Hazeltine, Department of Psychological and Brain Sciences, DeLTA Center, University of Iowa; Deborah K. Reed, Iowa Reading Research Center, College of Education, University of Iowa; Bob McMurray, Department of Psychological and Brain Sciences, and Department of Communication Sciences and Disorders, DeLTA Center, University of Iowa.

We thank Claire Goodwin and Patty Benzing for assistance with data collection; Amy VanHoosier and Eric Soride for assistance with the computer programming. We would like to thank Ariel Aloe for consultation on the statistical approach. This study was funded by an IES SBIR award to Carolyn Brown (Foundations in Learning, Inc.; Award S15C0023).

Correspondence concerning this article should be addressed to Tanja C. Roembke, Department of Psychological and Brain Sciences, DeLTA Center, University of Iowa, W311 SSH, Iowa City, IA 52242. E-mail: tanja-roembke@uiowa.edu

When children first learn to read an alphabetic language, it can be laborious. During the initial phases of phonics instruction, children are taught to sound out words, slowly producing each phoneme before fusing them to form a word.[1] With practice and instruction, children recognize known words by sight and decode new words increasingly rapidly (Ehri, 1995; Ehri & Wilce, 1983; Juel, 1983). By adolescence, many skilled readers can understand over a hundred words per minute in grade-appropriate texts. Although these gains in the speed of word reading are impressive, this is only part of skilled reading, as skilled reading requires a constellation of skills, including decoding, fluency, comprehension, and spelling. However, given that reading is a multifaceted process, how much does the rapidity of single word recognition matter?

In the Simple View of Reading, comprehension is built upon an individual's word recognition and language comprehension abilities (Gough & Tunmer, 1986). Children gradually learn to efficiently access their mental lexicon via written words, and this increasingly enables better developed spoken language comprehension (e.g., sentence processing, semantics) to support text compehension. The development of rapid single word recognition may be critical for skilled reading, as its emergence can gradually enable children to deploy more advanced, higher level oral language skills to reading.

Cognitive models of word recognition implicitly support the Simple View, emphasizing mechanistic routes by which efficient word recognition is achieved. Information processing models like the Dual Route Cascade model (Coltheart, Rastle, Perry, Langdon, & Ziegler, 2001; Roberts, Rastle, Coltheart, & Besner, 2003) assume word recognition becomes more rapid when children add individual orthographic word forms to their written lexicon (e.g., "memorizing" them; Share, 1995, 2008; Tamura, Castles, & Nation, 2017). Connectionist models (Plaut, McClelland, & Seidenberg, 1995; Seidenberg & McClelland, 1989) posit multiple pathways of connections between orthographic, phonological, and semantic representations, which can be tuned to support more efficient word recognition. Thus, understanding how efficiency is achieved could help clarify these models. Moreover, although these models are consistent with the Simple View of Reading, they do not speak to how differences in word recognition relate to educational outcomes like decoding or fluency, a crucial step in bridging cognitive models to real-world outcomes.

Studies confirm that word recognition accuracy is a critical predictor of outcomes for young readers (Catts, Fey, Zhang, & Tomblin, 1999; Cutting & Scarborough, 2006), even after controlling for listening comprehension (Curtis, 1980). There is also evidence that the speed of word recognition accounts for unique variance over and above these other skills across the elementary and high school years (Cutting & Scarborough, 2006). Such findings parallel the interest in measures such as Rapid Automatized Naming (RAN) as predictors of reading outcomes (Compton, 2003; Georgiou, Parrila, Cui, & Papadopoulos, 2013; Kirby et al., 2010; Lervåg & Hulme, 2009; Moll, Fussenegger, Willburger, & Landerl, 2009; Nag & Snowling, 2012; Pan et al., 2011; Powell, Stainthorp, Stuart, Garwood, & Quinlan, 2007; Wijayathilake & Parrila, 2013). Despite this interest in the speed of word recognition, there is still a need for further work to investigate the cognitive underpinnings of rapid lexical processing. Critically, understanding how lexical entries are quickly accessed may be crucial for accurately measuring word recognition.

Speed is often conceptualized in terms of automaticity. Automatic processes occur immediately and without effort or awareness (Logan, 1997). Automaticity can be straightforward to quantify for skills acquired in the laboratory, but it is more difficult to measure for skills acquired over months or years (while a child is also developing a variety of other skills). In such contexts, it is challenging to divorce the automaticity of a specific skill from increased mastery of the relevant knowledge (e.g., of the letters or words) or global changes in factors such as overall speed of processing (Kail & Hall, 1994).

The present study applies ideas from cognitive psychology on the automaticity of word recognition to education research. Our goals were to develop a new measure of automaticity and relate it to reading outcome measures such as fluency. This may be crucial not only for understanding the cognitive basis of this important reading achievement, but also for clarifying cognitive models on how automatic word recognition is achieved.

Specifically, we attempted to isolate automaticity from reading knowledge and from global processing speed. We did this with a backward masking paradigm. In a typical backward masking task, a stimulus is presented for a short amount of time and then covered, thus requiring automatic activation of the original stimulus for recognition. This avoids confounds with general processing speed, as the primary measure is accuracy and the response is untimed. Moreover, assessing performance on trials in which automaticity is not required (unmasked) can help control for knowledge of the relevant letters and words.

In addition to this primary goal, word recognition involves mappings between multiple components of reading such as phonology, semantics, and orthography (Seidenberg, 2005). Consequently, there may be multiple ways by which automatic word recognition can be achieved. Thus, we assessed automaticity using several new tasks with both words and nonwords. Including nonwords allowed us to force participants to rely on connections between phonology and orthography only. By examining multiple types of tasks, we could investigate multiple pathways by which words are accessed automatically, and whether any are uniquely predictive of outcomes.

Measuring Automaticity: Backward Masking

Automaticity in word recognition is not easily measured for three reasons. First, automaticity can be difficult to divorce from knowledge of the letters or words. Children who encounter a word that is not well-known may be slow because they have not mastered all of the component grapheme-phoneme-correspondence (GPC) regularities, and not because of a lack of automaticity. Indeed, this is one of the reasons why RAN measures focus exclusively on overlearned words such as color or number terms (Denckla & Rudel, 1974, 1976).

[1] Even though orthographic transparency differs across alphabetic languages, it is clear that phonological skills are universally important for reading acquisition (Goswami, Gombert, & de Barrera, 1998). Thus, although children may differ in how quickly they improve in word recognition (depending on the transparency of their native language), they often start off with decoding the letter-to-sound-mappings.

Second, older children as well as children with better nonverbal and language skills relative to age-matched controls are simply faster on virtually any language, reading, or nonverbal task (e.g., Kail, 2000; Miller, Kail, Leonard, & Tomblin, 2001). As a result, the consistently observed correlations between faster word recognition and reading outcomes could simply derive from differences in speed of processing. In this vein, Kail and Hall (1994) showed that the global development of speed of processing—not automaticity of a specific reading skill—drove differences in RAN and this was linked to reading comprehension. However, automaticity was not directly assessed, but rather was conceptualized as a function of experience. Thus, it is unclear to what extent automaticity in word recognition or more general increases in processing skill contributed to changes in reading outcomes.

Third, measures based on reaction time (RT) are the sum of many component processes. These include reading, but also processes such as response selection, motor planning, and motor speed (Luce, 1986). In addition, less relevant factors such as the child's general orientation to respond carefully or quickly may impact task performance (Förster, Higgins, & Bianco, 2003). The latter makes it particularly difficult to isolate automaticity in reading with RT alone.

One way to overcome these difficulties is the use of backward masking. In masking paradigms, words are presented briefly (e.g., 100 ms) and covered by a visual mask (e.g., #####). The visual mask simultaneously ends any bottom-up input and fills purely visual short-term buffers. Thus, input traces in sensory stores are unavailable to higher-level processing stages (cf., Breitmeyer & Ogmen, 2000; Turvey, 1973). Therefore, phonological or semantic codes must be maintained in the absence of confirmatory orthographic input. As we use masking here, participants select one of several response options depending on the task (e.g., the rhyme of the word that was presented, a picture corresponding to the target word, etc.) after the mask. Participants' responses are untimed, and the primary measure is accuracy. To achieve suitably strong activation of the correct word to generate a response, activation must be spread automatically from orthographic to semantic or phonological representations before the mask appears. Although backward masking requires rapid activation of semantic or phonological codes, it places no temporal pressure on later components of the system: As long as the correct phonological or semantic codes are active, the speed of response selection, planning, and execution is irrelevant (see Urry, Burns, & Baetu, 2015, for a similar argument for accuracy-based measures in sequence learning).

Backward masking has a long history in theoretical investigations of word recognition. For instance, it has been used to investigate processing differences between words and nonwords (Reicher, 1969; Wheeler, 1970). A seminal study by Forster and Davis (1984) used backward masking in combination with priming. They found that nonwords facilitated lexical decisions if they were similar to the target word; whereas, similar words impeded lexical decisions. This interaction suggests that in the brief period (60 ms) before the mask, skilled readers could access their lexica—otherwise words would have behaved like nonwords (Carreiras, Perea, & Grainger, 1997; Davis & Lupker, 2006).

Masking's importance in experimental work is well-documented. Here, we investigated whether it could be used as a diagnostic of individual differences in automaticity and a predictor of outcomes. In this regard, masking solves several problems with other measures of automatic word recognition.

First, even though masking imposes the need for automaticity, the response is untimed. This untimed response isolates the contribution of word recognition to RT from the time it takes for processes like decision making, response planning, or motor execution. Relatedly, speeded tasks typically present a tradeoff between speed and accuracy: Children can choose to respond quickly at the cost of accuracy. Individual differences in accuracy and response speed could thus impact general RT measures of automatic word recognition (Förster et al., 2003). In contrast, masking removes the pressure to respond as quickly as possible, so children can be more careful. However, accurate responses still require rapid activation of the reading system.

Second, the flexibility of the task enables potentially richer measures: Backward masking can be used with both familiar words and unfamiliar nonwords, and with several dependent measures (e.g., accuracy, effect of priming; Frost & Yogev, 2001; Perfetti & Bell, 1991; Perfetti, Bell, & Delaney, 1988). Crucially, the same or matched items can easily be assessed in both masked and unmasked form. The latter can provide a useful control to estimate the child's knowledge of the word and relevant GPC regularities when there is no requirement to access this knowledge automatically (cf. Stanovich, 1980).

Finally, masking takes away the sustained bottom-up input, making experiments potentially more sensitive to differences between children: Under typical conditions, activation and competition between words are continuously affected by the visual input. For example, when CAT is read, CAP briefly competes. However, as long as the T is visible, the correct lexical representation will be most active. Thus, differences in competition across participants may only appear as small differences in RTs. By removing the bottom-up information (the T), competition can unfold independently of this input. As a result, differences may be large enough to emerge even in the final decision (e.g., in measures of accuracy).

Multiple Loci of Automatic Word Recognition

Although masking may offer a unique way to assess automaticity, automaticity in word recognition may also be divorced from global processing speed by considering the types of subprocesses that may be automatic. If automaticity can be pinned to particular subcomponents of written word recognition, a stronger case would be made for automaticity as a component of word recognition rather than as a general indicator of processing speed (a form of discriminant validity). A critical way to evaluate this would be (a) to measure whether automaticity can even be identified within specific subprocesses in word recognition; and (b) if so, to assess whether the degree of automaticity in these subprocesses differentially predicts reading outcomes.

Traditional models of reading assume that decoding (mapping sounds to letters) is slow. For example, the Dual Route Cascade (DRC) model (Coltheart et al., 2001; Roberts et al., 2003) assumes decoding is a serial process, in which letters are individually mapped to pronunciations, one by one. Automaticity in word recognition is then achieved when children learn or memorize specific words and access them directly without decoding (Share, 1995, 2008; Tamura et al., 2017). This predicts that automaticity can only be achieved for well-learned, known words.

Connectionist models also posit two ways by which the meaning of a word can be accessed (see Figure 1; Plaut et al., 1995; Seidenberg & McClelland, 1989). Readers can access meaning from spelling directly (orthography to semantics, $O \rightarrow S$) or indirectly by using the systematic mapping from orthography to phonology (decoding) and then the well-learned phonology-to-semantics-mapping (spoken word recognition; $P \rightarrow S$). Both the $O \rightarrow S$ and $O \rightarrow P \rightarrow S$ pathways are implemented over learnable networks of connections, which transmit activation more or less efficiently depending on how well these connections are organized.

In the multipathway-framework, automaticity could be conceptualized in two ways. First, it may reflect the relative importance of pathways. During typical reading, both phonological and semantic information is activated by the orthographic input. Thus, readers use both the indirect $O \rightarrow P \rightarrow S$ and the direct $O \rightarrow S$ pathways. However, the degree to which word recognition in a given context relies on each pathway may be influenced by task demands, the familiarity of specific words or GPC regularities, and the coherence of each pathway (Chace, Rayner, & Well, 2005).[2] Analogous to traditional models of reading (such as the DRC model), the indirect $O \rightarrow P \rightarrow S$ pathway may be too slow for rapid reading. In this case, faster word recognition may indicate dominance of the direct $O \rightarrow S$ pathway (Harm, Seidenberg, Macdonald, Thornton, & Zevin, 2004). Hence, the ability to directly activate semantic information from orthography will better predict outcomes than the ability to activate it from phonology. Pedagogically, this conceptualization of reading suggests automaticity is only achieved through large amounts of practice enabling children to directly map written words to lexical or semantic representations.

Second, automaticity could be achievable in both pathways to various degrees or even specific subsets of associations within a pathway (e.g., for a class of GPC regularities within $O \rightarrow P$). Consequently, a child could be more automatic in one mapping (e.g., $O \rightarrow S$) than the other ($O \rightarrow P \rightarrow S$), and use the pathways simultaneously and interchangeably, depending on the situation. This account also predicts that even well-known words could vary in automaticity, or that two children who both know a given word could differ in their degree of automaticity. More importantly, it predicts that automaticity also can be achieved in nonwords, which must rely on the $O \rightarrow P$ pathway (Hogaboam & Perfetti, 1978; Perfetti & Hogaboam, 1975). It follows that measures of automaticity in each pathway should predict reading outcomes, though not necessarily the same outcomes.

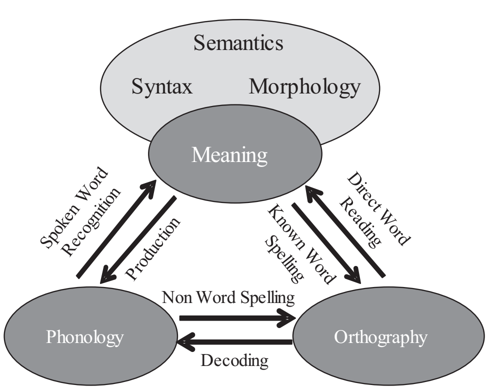

Figure 1. Depiction of the theoretical relationship between orthography, phonology and meaning. Adapted from "A Distributed, Developmental Model of Word Recognition and Naming.," by M. S. Seidenberg and J. L. McClelland, 1998, *Psychological Review, 96,* p. 526. Copyright 1998 by the American Psychological Association.

What Outcomes Are Related to Automaticity?

Automatic word recognition may predict some components of reading success but not others. Such a differential pattern of predictive relationships may offer a second form of discriminant validity for identifiying a unique role of automaticity. Reading automaticity theory (LaBerge & Samuels, 1974) as well as the verbal efficiency theory (Perfetti, 1985) assigns automaticity a key role in reading development (Klauda & Guthrie, 2008; Kuhn, Schwanenflugel, Meisinger, Levy, & Rasinski, 2010; Rasinski et al., 2005; Schwanenflugel et al., 2006). These theories argue that increased automaticity in written word recognition is crucial for fluent reading (the ability to read connected text accurately, quickly and, in the case of oral reading, with expression; National Reading Panel, 2000). Higher levels of fluency free resources for comprehension. That is, as word recognition becomes faster with practice, children can recognize larger units in a single fixation, such as phrases stored in the lexicon, freeing attention and working memory capacity for reading comprehension (Walczyk, 2000).

Such models make sense when considering children's prereading skillsets. For most school-age children, spoken language comprehension is fairly automatic. As a result, when reading is fast and fluent, written text becomes another way to access better-learned spoken language (as in the Simple View of Reading). In contrast, if fluency is poor, it may be too laborious or memory-intensive to engage these processes, leading to poor comprehension. Consistent with this explanation, listening comprehension and decoding account for a majority of variance in reading comprehension (Catts et al., 1999; Cutting & Scarborough, 2006; McCardle, Scarborough, & Catts, 2001; Nation & Snowling, 1998, 2000; Scarborough, 1990).

However, there are surprisingly few studies that directly investigate the relationship between automaticity and outcomes (though see Schwanenflugel et al., 2006). A number of studies have examined the relationship between fluency and comprehension and other outcomes (Cutting & Scarborough, 2006; Jenkins, Fuchs, van den Broek, Espin, & Deno, 2003; Kim, 2015; Klauda & Guthrie, 2008; Nunes, Bryant, & Barros, 2012). However, fluency is a constellation of skills and an important outcome in its own right. Other studies have examined speed of processing of well-known items such as letters (e.g., Joshi & Aaron, 2000).

One way to investigate the importance of automaticity is to target a population, such as older students, that may struggle with automatic word recognition despite adequate knowledge of GPC regularities. By middle-school, pedagogical emphasis is largely on text comprehension, not word-level skills. A review of current

[2] This is not to say that phonology is not used at all under some circumstances; ample data suggest phonology is activated rapidly during reading, even in highly skilled adult readers (e.g., Leinenger, 2014). Rather, successful identification of a word's meaning may be influenced more (or less) by one pathway or the other.

literacy standards (e.g., National Governors Association Center for Best Practices & Council of Chief State School Officers, 2010) generally indicates no expectations for building phonics or fluency skills beyond Grade 5. At this point, continued growth in reading ability is assumed to rest on the development of more advanced skills such as comprehension or inference (Cain, Oakhill, Barnes, & Bryant, 2001).

This strategy may be appropriate for successful readers. However, roughly half of struggling middle-school readers show deficits in basic word recognition, decoding and fluency (Cirino et al., 2013; Hock et al., 2009). These children have often been exposed to the relevant knowledge (e.g., GPC regularities, sight words), raising the possibility that some of their deficits lie in the ability to employ this knowledge automatically. Thus, below-average middle-school students, a chronically understudied population in basic research, may be well-suited for testing deficits in automaticity. Moreover, an emphasis on automaticity may offer a useful target of intervention for those middle-school readers who still struggle with word-level skills.

The Present Study

We asked how automaticity in two reading pathways relates to several outcomes in typical-to-low performing middle-school students. Three computer-based tasks were developed that use backward masking to assess automaticity. Tasks were designed to emphasize different pathways of the visual word recognition system. One emphasized semantics by asking children to select a picture for a given word, and two included nonword stimuli to more directly isolate O → P processing in the absence of a semantic or lexical representation for the item. Each task was run both with and without backward masking to disentangle students' knowledge of the GPC system and the words from their ability to deploy this knowledge automatically. The relationship between performance on these three tasks and standardized measures of reading comprehension, fluency, and decoding was explored in a hierarchical analysis that controlled for language ability and performance on unmasked version of the same task. This design addressed two research questions:

1. Is automatic word recognition correlated with reading outcomes (decoding, fluency, comprehension), over and above oral language ability and general differences in reading knowledge (the unmasked versions of the tasks)?

2. If automaticity does predict reading outcomes, can it be observed in each pathway (O → P, O → S) within the reading system? For this purpose, we replicated the previous analysis with nonword trials (if available) to estimate the extent to which automaticity may exist in the O → P pathway.

The three experimental tasks assessed both knowledge and automaticity along both reading pathways. These pathways cannot be truly isolated, as readers likely use a combination to solve most tasks. Instead, our goal was to develop tasks that change the degree to which either the O → S or the O → P pathways contribute to word recognition.

To emphasize O → S processing, we used a picture matching task (Find the Picture). Participants read a word and selected its picture from a screen containing four pictures. It was inspired by oral vocabulary tests like the Peabody Picture Vocabulary Test (PPVT). As the target word was presented orthographically, it required students to map the written word onto one of the depicted concepts. Competitors were orthographically similar to the target.

To emphasize the O → P mapping, we used a rhyme task (Find the Rhyme) in which participants read a word or nonword and selected which of eight printed competitors rhymed with it. As this stressed the phonological form of the word, it was expected to highlight the O → P pathway. On nonword trials, this was likely the only pathway available. Find the Rhyme was built on the basis of findings that identifying rhymes is an important phonological skill that is predictive of later reading development (e.g., Bryant, MacLean, Bradley, & Crossland, 1990).

Finally, in the word verification task (Verify), participants heard a word and saw a written word that could either match or mismatch it. Although this is primarily a phonological task, the emphasis on whole words was expected to evenly tap both pathways, though again on half of the trials nonwords were used to isolate the O → P pathway. This task was developed based on matching tasks that test how auditory and orthographic information are compared.

Absolute performance on these tasks cannot be compared as they used different numbers of response options. For instance, Find the Picture used only four options because more would have overcrowded the screen and exceeded the capacity of visual working memory (Luck & Vogel, 1997). Find the Rhyme used eight choices. This larger number of response options was necessary to capture a variety of similarity relationships to the target, which was essential to focus the child's attention on the entire string. Finally, Verify had only two options (match/mismatch). These differences would be problematic if our goal was to compare tasks (e.g., to profile the strengths and weaknesses across pathways). However, our goal was correlational, asking whether automaticity in each pathway was a better predictor of outcomes. Thus, it was more important that tasks were appropriate for the question than directly comparable.

Visual backward masking was used for each task to assess the degree of automatic text processing. A reduction in accuracy across all three tasks was predicted with masked stimuli. This study's primary question was whether masked accuracy uniquely predicted outcomes over and above unmasked accuracy on the same task. In this context, accuracy on unmasked items can be seen as a marker of what the student knows; whereas, accuracy on masked items assesses whether students process these items automatically.

Participants were middle-school students falling below the 60th percentile on a standardized test of reading comprehension. In middle school, students should be adequate at decoding (i.e., their knowledge of the GPC rules should be well-developed) but may still lack automaticity to activate words in-the-moment. Thus, investigating automaticity in word-level reading mappings may help reveal how best to assess and remediate this group.

Method

Participants

Participants were 58 students ($n_{\text{female}} = 32$, $n_{\text{male}} = 26$) at a middle-school in Iowa. Thirty-one students were in Grade 7 (53%) and 27 were in Grade 8 (47%). Their mean age was 13.2 years

(*SD* = 8 months). All were native speakers of English who self-reported normal hearing. Five students (9%) qualified for individualized education programs (IEPs), and one student (2%) qualified for a 504 plan. None of the students had an intellectual disability. It was not possible to attain additional information on students' disability status. Forty-three of the students qualified for free or reduced price lunch (74%), indicating a lower socioeconomic status (SES). Thirty-nine (67%) identified as White, 17 (29%) as African American, one (2%) as American Indian or Alaskan Native, and one (2%) as Asian. None of the participants were classified as English language learners. Recruitment targeted students between the 10th and 60th percentile on their previous year's reading subtest of the Iowa Assessment, which measures silent reading comprehension. Two students were excluded, as they completed less than half of the trials, leaving 56 students in the final analysis. Consent forms were completed by parents or guardians in accord with an institutional review board (IRB) approved protocol. In addition, students completed a group administered assent protocol, where they agreed to participate via raising their hand and saying "yes" aloud. The study was approved by the IRB.

Standardized Test Battery

Scores from six standardized assessments were collected; these assessments are commercially available and have extensive psychometric data.

Decoding. Word-level reading was assessed with two subtests of the Woodcock Johnson Reading Mastery Test, 3rd ed. (WRMT; Woodcock, McGrew, & Mather, 2001). In Word Identification (WRMT_ID), students read familiar words of increasing difficulty from a test easel; accuracy is scored. In the Word Attack (WRMT_AT), they read aloud low frequency and nonsense words that increase in difficulty, with accuracy as the measure. In both, responses are untimed. Test–retest reliability for both at Grade 8 is reported by the developers as 0.98 and 0.90, respectively (Woodcock et al., 2001).

Oral fluency. Oral fluency was assessed with the Texas Middle School Fluency Assessment (TMSFA), a timed test of oral passage reading rate and accuracy (Francis, Barth, Cirino, Reed, & Fletcher, 2010). Participants read short passages with the instruction to do their "best reading." Scores are based on the number of words read correctly. Developers reported that all passages were equated, and the mean intercorrelation on the five passages across testing points from 0.86 to 0.98 in 6th- to 8th-grade students (Texas Education Agency, University of Houston, and The University of Texas System, 2008).

Silent fluency. The Test of Silent Contextual Reading Fluency (TOSCRF) was used to assess silent reading fluency. The TOSCRF is a timed task in which students identify word boundaries within sentences and paragraphs without spaces (Hammill, Wiederholt, & Allen, 2006). The number of correct boundaries is scored. The TOSCRF has high test–retest reliability ($r = .89$) with middle-school students (Hammill, Wiederholt, & Allen, 2014).

Comprehension. Reading comprehension was assessed with the comprehension subtest of the Gates-MacGinitie Reading Test, 4th ed. (GMRT). Students have 35 min to read a series of short passages and respond to multiple-choice questions (MacGinitie, MacGinitie, Maria, & Dreyer, 2000). Questions tap both literal and inferential comprehension. Developers of the GMRT report high test-retest reliabilities ($r = .85$ and above; MacGinitie et al., 2000).

Finally, the untimed PPVT was used to assess oral language. Students heard an orally presented word and identified the picture that matched from an array of four pictures (Dunn & Dunn, 2007). For children 11–14, the PPVT-4 has excellent test–retest reliability ($r = .93$). The PPVT (vocabulary) was treated as a covariate (not an outcome) because previous research has shown that vocabulary may contribute heavily to reading outcomes (Duff, Reen, Plunkett, & Nation, 2015; Jenkins et al., 2003; Oslund, Clemens, Simmons, Smith, & Simmons, 2016; Wegener et al., 2017).

WRMT, TOSCRF, and PPVT were reported as standard scores. For GMRT, standard scores could not be obtained directly but were converted from percentiles. Standard scores are not available for the TMSFA. Instead, passage scores were converted to an equated score (taking passage difficulty into account), and equated scores were averaged.

Procedures. Students first completed the group-administered paper-and-pencil tests, the GMRT and TOSCRF. The GMRT was administered by the lead tester with the assistance of two additional testers at the beginning of each testing session in a classroom with approximately 20–30 students. Subsequently, participants were separated into groups of four to complete the TOSCRF. After group testing, participants were assigned to one of four testers for individual tests. As testing was completed within the same classroom, testers spread out to its four corners to minimize noise levels. Each tester started with a different test and then followed the same order: PPVT, WRMT_ID, WRMT_AT, and TMSFA. This procedure assured that no students completed the same assessment at the same time and controlled for order effects.

The four testers (three females) were extensively trained by the research team on standardized assessments. Assessments were scored by two testers. In the rare event of a discrepancy, the lead tester checked the document a third time and confirmed the correct score.

Experimental Measure

All the experimental tasks were delivered via a custom program over the Internet. They were part of a development version of the Iowa Assessment of Skills and Knowledge for Reading, a computer-based reading assessment for struggling middle-school readers. The data presented here are part of a larger set of trials that were designed to complete four independent experiments on middle-school word recognition simultaneously. Trials for all four experiments were interleaved as part of a single continuous program (from the participant's perspective). In a sense, the trials for Experiment 2 (e.g.) served as filler trials for Experiment 1. Each experiment comprised multiple 20-trial blocks, and a given block contained trials for only one experiment. However, blocks were interleaved such that participants would complete trials for all four experiments in close proximity and in a single session. The trials not used in this study addressed independent questions such as how word length, number of syllables and a stimulus' stress pattern influence performance.

Procedures. Participants completed 1,380 trials over 4 days as part of the larger sample, approximately 345 trials per day. Of those trials, 600 were designed for the present study and are analyzed in this article. The other 780 were used for other experiments addressing different goals from the present study. On each day, students were tested for approximately 45 min. Trials for this specific study were interspersed throughout the four days of data collection.

On each day, students logged into the experimental program and saw a selection screen with icons for a small number of tasks. Each selection screen provided a complete set of tasks, but students could complete them in any order. Masked and unmasked versions of the same task (e.g., Verify and Verify Masked) never appeared on the same screen. Whether a student did the masked or unmasked version of a particular task first was counterbalanced. After completing all task versions on a selection screen, students moved on to a new screen. Successive screens were a lateral progression and not an increase in difficulty level. Each iteration of a task within a selection screen included 20 trials, and each task appeared on multiple screens to achieve more trials. Blocks of trials for this experiment were interleaved across all screens.

Design. Three base tasks were used: Find the Picture, Find the Rhyme, and Verify. Each featured an orthographically presented target stimulus in lower cases, followed by a response. Each task had an unmasked and masked version. In the unmasked version, participants had as much time as needed to view the stimulus and respond. In the masked version, target stimuli disappeared and were replaced with a mask after 90 ms. The mask was always eight repeated hashtags (########; Figure 2D), exceeding the number of letters of the longest stimulus. The viewing period was determined based on pilot testing of 55 students from the same middle-school, using similar tasks with various viewing periods. These data and focus groups with the students were used to find an appropriate masking interval.

A trial always started 1,600 ms after the previous one. There was no fixation cue or "go button" that had to be clicked before a new trial. The assignment of items to masking conditions was counterbalanced such that words that appeared in the masked version of a task for one student were in the unmasked version for others, and vice versa. In addition, there were two lists for the Verify task to balance whether the auditory word matched or mismatched the target visual word (see below). Thus, students were randomly assigned to four randomization lists.

Students completed 240 trials (120 masked) in Find the Rhyme and Verify each. Half of the stimuli were words, and the other half nonwords (balanced within block and across masked and unmasked versions). Find the Picture used only 120 trials as it could only test real words. On average, participants completed 576 (standard deviation ± 57) trials. Thirty participants completed the full 600 trials. Because of how the different tasks were distributed over the days, participants were more likely to miss trials in the Find the Picture task version (M = 49/60 trials ± 13) than all other task versions (Find the Picture masked M = 59/60 trials ± 5; Find the Rhyme M = 117/120 trials ± 10; Find the Rhyme masked M = 117/120 trials ± 12; Verify M = 117/120 trials ± 10; Verify masked M = 117/120 trials ± 12).

To assess reliability, Cronbach's alpha was calculated separately for each unmasked and masked task version. As there were four randomization lists, alpha was calculated separately for each list and averaged. For children who did not complete all of the trials for each task, we estimated the number of correct responses they would have had for all trials based on the proportion correct for those they completed. To do so, we multiplied the proportion correct (on completed trials) by the total number of items for that task. For participants who did not miss any trials, this was the same total as the number of correct trials. For participants who missed a few trials, this method essentially filled in missing trials with the average of the other items. These estimated values were only used

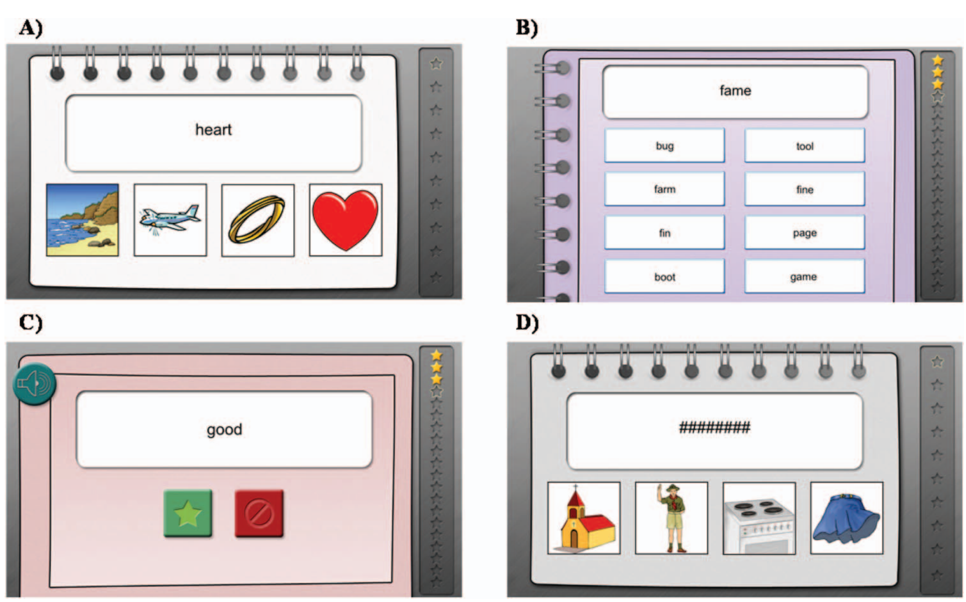

Figure 2. Screenshots of the three tasks. Panels A–C show the screen before masking and Panel D represents an example of the masked version. A: Find the Picture; B: Find the Rhyme; C: Verify; and D: Find the Picture with mask. Yellow stars on the side indicate completed trials. See the online article for the color version of this figure.

for computing Alpha, not for any other analyses. For the unmasked version of Find the Picture, average Cronbach's alpha was 0.90; its masked version was similarly reliable ($\alpha = .92$). Find the Rhyme also had excellent reliability (unmasked: $\alpha = .98$; masked: $\alpha = .97$) as did Verify (unmasked: $\alpha = .98$; masked: $\alpha = .90$. Thus, experimental tasks showed excellent internal consistency.

Tasks. Participants received audio instructions at the beginning of each task version. Trials were completed when the participant selected a response. No feedback was given based on the response, but participants received occasional audio encouragement (e.g., "Keep up the good work!") between trials. An experimenter was present during sessions to assist with computer problems, answer student questions, and monitor student attention.

In Find the Picture, participants saw a target word and four picture response options arranged horizontally. Pictures represented either the target word or one of three foils (see Figure 2A). Foils were drawn from an extensive database of possible items. Foils were chosen to be orthographic or phonological competitors of the target word (e.g., COMB as a competitor for COAST). Semantic competitors (e.g., BEACH as a competitor for COAST) were never included to avoid confusion when identifying the semantic representation of a word.

In Find the Rhyme, participants saw a target word with eight written response options, arranged in two columns of four (see Figure 2B). One of the options was a correct response (a rhyming word or nonword); the other seven were foils. Approximately 60% of the correct responses used the same orthographic vowel or vowel string as the target word (e.g., target: MAIL; rhyme: SAIL), the remaining trials used a different letter string with the same pronunciation (e.g., target: MAIL; rhyme: SALE). These were included to force participants to encode stimuli phonologically. Thus, the task could not be solved by looking for the matching letter string. The seven foils followed a characteristic structure. Two began with the same consonant(s) and vowel (based on pronunciation) as the target word (e.g., for MAIL, MAZE and MAIN).[3] Two matched on the beginning and ending consonant(s), (e.g., MULE and MEAL). Two matched the target word's vowel only and often matched a consonant with the rhyme (e.g., SAME and SANE). The last one matched on any other dimension (e.g., RAIN). It was not always possible to find a word for every slot in this scheme due to limitations in the English vocabulary. For word stimuli, all of the response options were real words; for nonword stimuli, all of the responses were nonwords. Foil order was randomly determined on each trial.

In Verify (Figure 2C), participants saw a word or nonword and simultaneously heard a spoken item. They then indicated whether the auditory stimulus matched the written item by clicking a red button for "no" and a green one for "yes." They could hear the auditory stimulus again via a button if desired. Auditory and visual stimuli matched on 50% of trials. Foils for nonmatching trials were phonologically and orthographically similar to the target with onset, offset, and other types of matches possible. For example, the competitor for COAST could be COACH (onset competitor), TOAST (offset competitor) or CLOSED (other). Offset competitors would always be rhymes (TOAST), and sometimes also rimes (GHOST). Two lists of match and nonmatch items were counterbalanced across participants. Items in List 1 for which the auditory and orthographic stimuli did not match were presented in the match condition in List 2.

Items. Items were chosen to be words that were likely known by middle-school students and had appropriate content. To do so, words were evaluated by a group of six people (the four authors plus two collaborators from Foundations in Learning). Five were native speakers of English; one was a developmental psychologist (Bob McMurray); one was a former secondary teacher and current education researcher (Deborah K. Reed); and one was a curriculum developer with experience in middle-school.

All items were monosyllabic. Monosyllabic words were chosen for several reasons: First, they allowed us to control a number of item factors, such as the prevalence of vowel types across words (see below). Second, one of our goals was to control the similarity of orthographic competitors as much as possible; this is facilitated by the use of monosyllabic words, as they tend to have a higher number of competitors than multisyllabic words.

Items were systematically selected to sample a wide range of forms, and we controlled several stimulus characteristics across tasks and conditions. Half of the items included a consonant cluster or digraph; the other half had only simple consonant-vowel-consonant frames. In addition, items were balanced by vowel GPC regularities, with each class including multiple vowel strings. We included short vowels (e.g., A as in CAT), long vowels (e.g., I_E as in LICE), secondary short and long vowels (e.g., I as in BLIND, O_E as in LOVE), vowel digraphs (e.g., EE as in BREED), secondary or exception digraph vowels (e.g., EA as in BREAD), and diphthong and R-controlled vowels (e.g., OI as in BOIL, AR as in BARK). An approximately equal number of words and nonwords was included for each GPC class, other than the two exception classes where nonwords were not possible. Nonwords were phonologically legal pseudowords in English (e.g., LOIF, GNAST). They were constructed to follow similar phonotactic sequences as existing words and validated by separate experimenters. Nonwords did not include morpheme-like substrings. Roughly equal numbers of items of each GPC class and consonant difficulty were assigned to the masked and unmasked versions of all tasks, and to the match and mismatch conditions in Verify.

Auditory stimuli. Stimuli were recorded by a phonetically trained female talker. Stimuli were spoken clearly at a slow rate in a neutral carrier sentence ("S/he said . . .") to ensure uniform prosody. Recordings were conducted in a sound-attenuated room using a Kay Elemetrics CSL 4300b (Kay Elemetrics Corp., Lincoln Park, NJ) at 44,100-Hz. Stimuli were edited to remove the carrier sentence and to eliminate any distracting articulations (e.g., jaw clicks, breaths). Fifty ms of silence were appended to the beginning and end of each word. The final stimuli were amplitude normalized using Praat (Boersma & Weenink, 2016).

Visual stimuli. Pictures in the Find the Picture task were colored line drawings developed using a standard lab protocol (McMurray, Samelson, Lee, & Tomblin, 2010). For each item, 8–10 candidates were downloaded from a commercial clip art database. These were viewed by a focus group of undergraduate and graduate students to select the most representative one for each word. Selected pictures were edited to ensure uniform style as well as prototypical colors and orientations, and to remove any distract-

[3] For purposes of matching foils, the silent E in long vowels such as MULE was treated as part of the vowel.

ing elements. Final pictures were approved by Bob McMurray who has extensive experience with similar tasks.

Results

Our analysis started with three preliminary methodological issues: an overview of outcome measures to characterize the sample, an analysis of accuracy to document that masking had the expected effect, and an overview of the intercorrelations among tasks. Next, we turned to our two research questions. The first question was addressed with hierarchical regressions determining whether masked performance accounts for variance in outcomes over and above unmasked performance. We also conducted a communality analysis to estimate the shared variance between unmasked and masked performance and their unique contributions. Subsequently, the second question was addressed with hierarchical regressions of only nonword trials to determine if automaticity could be detected in O → P processing.

Outcome Measures

The distributions of outcome measures were investigated for outliers, skewness, or other abnormalities. All variables were found to be suitable for the planned statistical analyses. Means and standard deviations of performance on the assessments are in Table 1. As our sample targeted middle-school students with average to below-average performance on reading comprehension, the descriptive results of the word-level measures (decoding: WRMT_ID, WRMT_AT; fluency: TMSFA, TOSCRF) confirmed the appropriateness of the sample for our research questions. Given the very high correlation between WRMT_ID and WRMT_AT (r = .614) and their conceptual overlap, these two scores were averaged to create an overall WRMT score (see Supplement S1 in the online supplemental material for correlations among all measures). To facilitate readability, we refer to assessments by the construct they assess (WRMT = decoding; GMRT = comprehension; TMSFA = oral fluency; TOSCRF = silent fluency; PPVT = vocabulary).

Effect of Masking

Accuracy was lower on masked than unmasked versions of all three tasks (see Figure 3). This decrement was particularly pro-

Table 1
Overview of Reading Outcomes

Variable	*M*	*SD*
WRMT_ID	92.30	13.01
WRMT_AT	90.46	14.72
TMSFA	128.41	26.84
TOSCRF	85.89	11.47
GMRT	95.24	9.65
PPVT	96.75	10.47

Note. All measures are given as standardized scores with the exception of TMSFA, which is reported as the average equated value of the number of words read correctly per minute. WRMT_ID = Woodcock Reading Mastery Word Identification; WRMT_AT = Woodcock Reading Mastery Word Attack; TMSFA = Texas Middle School Fluency Assessment; TOSCRF = Test of Silent Contextual Reading Fluency; GMRT = Gates-MacGinitie Reading Test; PPVT = Peabody Picture Vocabulary Test.

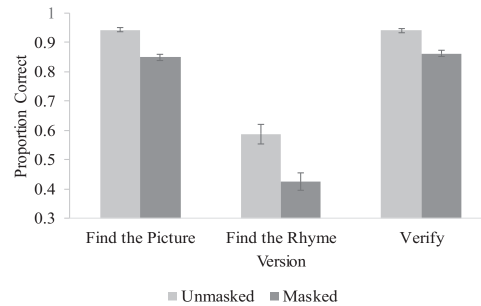

Figure 3. Proportion correct across the three tasks for both the masked and unmasked variant. Error bars indicate the standard error of the mean.

nounced and more variable in Find the Rhyme. However, chance performance was different for each task (Find the Picture: 25%; Find the Rhyme: 12.5%; Verify: 50%), making it difficult to make direct comparisons of absolute accuracy across them. This likely accounts, in part, for the lower performance in Find the Rhyme.

A series of 2 × 2 repeated-measures analyses of variance (ANOVAs) examined accuracy (empirical logit transformed) as a function of masking (masked or unmasked) and item-type (word or nonword), both within-subject factors. ANOVAs were conducted separately for Find the Rhyme and Verify. For Find the Picture, an ANOVA with only masking as a factor was conducted, as the task only includes real words. The main effect of masking was significant for all three tasks: Find the Picture: $F(1, 55) = 79.18$, $p <$ .001, $\eta_p^2 = 0.590$; Find the Rhyme: $F(1, 55) = 98.38$, $p < .001$, $\eta_p^2 = 0.641$; Verify: $F(1, 55) = 120.61$, $p < .001$, $\eta_p^2 = 0.687$. As predicted, performance was lower on masked than unmasked versions. In addition, the main effect of item-type as well as the interaction of masking and item-type were significant for Find the Rhyme (main effect: $F(1, 55) = 16.52$, $p < .001$ $\eta_p^2 = 0.231$; interaction: $F(1, 55) = 2.10$, $p = .037$, $\eta_p^2 = 0.037$) and Verify (main effect: $F(1, 55) = 56.88$, $p < .001$ $\eta_p^2 = 0.508$; interaction: $F(1, 55) = 20.82$, $p < .001$, $\eta_p^2 = 0.275$). The significant main effect of item-type was the result of lower performance for nonwords than words. The interaction indicated that participants showed a greater masking decrement for nonwords than for words.

Task Correlations

We examined the pattern of correlations among our experimental tasks as an exploratory analysis to understand how the different classes of tasks (Find the Picture, Find the Rhyme, Verify) relate, and to determine if performance on masked versions relates to each other. As Table 2 shows, after applying a conservative Bonferroni correction for multiple comparisons, most correlations were significant (all were significant at uncorrected alpha level). These correlations were not surprising given that tasks incorporated a similar set of items and had similar task demands. The exception was that most correlations with Find the Picture were not significant at corrected alpha levels, likely due to the lower number of trials included (though intriguingly, the masked variant did better).

Our small sample size and the large number of correlations (family wise error) prevents us from making strong inferential claims. However, in general, both versions of the same task

Table 2
Overview of Correlations Between Different Tasks

Task	1	2	3	4	5	6
1. Picture						
2. Picture Masked	.337					
3. Rhyme	.371	.521*				
4. Rhyme Masked	.304	.573*	.851*			
5. Verify	.558*	.312	.476*	.455*		
6. Verify Masked	.275	.561*	.567*	.636*	.574*	

Note. All uncorrected correlations were significant at $p < .05$.
* Correlations that remained significant after applying the Bonferroni correction ($p < .003$).

(masked and unmasked) were highly correlated. Moreover, across tasks, performance on masked versions of one task tended to be more highly correlated with other masked versions of other tasks than with unmasked versions of those other tasks.

Question 1: Does Automaticity of Word Recognition Predict Outcomes?

Hierarchical regression. Our primary question was whether masked performance accounted for additional variance in reading outcomes over and above what was accounted for by unmasked performance. Separate hierarchical regression analyses were run for each outcome, using a similar model. In these models, oral vocabulary was entered first because vocabulary contributes heavily to reading outcomes (Jenkins et al., 2003). On the second level, unmasked performance of a particular task (e.g., unmasked Find the Picture) was added to account for the relevant knowledge for completing that task. Finally, masked performance of the same task was added in the third level. Significance at each step was evaluated with ΔR^2 and the associated ΔF statistic, which conservatively applies all shared variance to the earlier step of the model.

Table 3 shows complete results of all analyses. Here, we summarize across individual analyses for the same outcome measure. For oral fluency (TMSFA), the unmasked versions of Find the Rhyme and Verify were significant predictors (unmasked Find the Rhyme: $\Delta R^2 = 0.10$; unmasked Verify: $\Delta R^2 = 0.096$). Crucially, the masked versions of Find the Picture and Find the Rhyme significantly predicted oral fluency over and above unmasked variants (masked Find the Picture: $\Delta R^2 = 0.08$; masked Find the Rhyme: $\Delta R^2 = 0.073$). Silent fluency (TOSCRF) showed a similar pattern. Both unmasked versions of Find the Rhyme and Verify significantly predicted silent fluency (Find the Rhyme: $\Delta R^2 = 0.213$; Verify: $\Delta R^2 = 0.109$). Most importantly, over and above unmasked tasks, silent fluency was significantly predicted by the masked version of Find the Picture ($\Delta R^2 = 0.069$).

For unspeeded word recognition and decoding (WRMT), vocabulary and unmasked versions of all three tasks accounted for significant variance (vocabulary: $\Delta R^2 = 0.176$; unmasked Find the Picture: $\Delta R^2 = 0.066$; unmasked Find the Rhyme: $\Delta R^2 = 0.258$; unmasked Verify: $\Delta R^2 = 0.223$). All of the masked versions captured extremely small amounts of variance (masked Find the Picture: $\Delta R^2 = 0.001$; masked Find the Rhyme: $\Delta R^2 < 0.001$; masked Verify: $\Delta R^2 = 0.002$), none of which were significant. These results suggest that unspeeded decoding and word recognition depend largely on knowledge, not automaticity. Similarly, reading comprehension (GMRT) was related to vocabulary (PPVT; $\Delta R^2 = 0.242$); neither unmasked nor the masked versions of the tasks accounted for significant variance.

These analyses also were conducted without the six students who either qualified for an IEP or 504 plan. The pattern of results was the same with several exceptions. For oral fluency, there was only a marginal effect of the masked version of the Verify task ($p = .081$), though this was still significant for Find the Rhyme ($p = .026$) and Find the Picture ($p = .016$). For silent fluency, the masked version of Find the Picture was no longer a significant predictor ($p = .145$), and the masked version of Verify was only marginal ($p = .093$). The masked version of Find the Rhyme remained a nonsignificant predictor as before ($p = .310$).

Communality analysis. In the previous analyses, unmasked performance subsumed shared variance with masked. Thus, it is unclear if unmasked performance also contributed unique variance to reading. Thus, we conducted a communality analysis to estimate the shared and unique variances for unmasked and masked versions of each task. For this purpose, a parallel analysis was conducted, in which vocabulary scores were entered in the first

Table 3
Overview of Regression Analyses

		1. Vocabulary			2. Unmasked			3. Masked		
Outcome	Task	*R*	ΔR^2	Sig. ΔF	*R*	ΔR^2	Sig. ΔF	*R*	ΔR^2	Sig. ΔF
Spoken Fluency	Picture	.182	.033	.18	.239	.024	—	.371	.08	.032*
	Rhyme				.365	.1	.016*	.454	.073	.033*
	Verify				.359	.096	.019*	.397	.029	.19
Silent Fluency	Picture	.127	.016	—	.263	.053	.087	.372	.069	.046*
	Rhyme				.478	.213	<.001*	.491	.012	—
	Verify				.354	.109	.013*	.392	.029	—
Decoding	Picture	.42	.176	.001*	.492	.066	.036*	.493	.001	—
	Rhyme				.659	.258	<.001*	.659	<.001	—
	Verify				.632	.223	<.001*	.634	.002	—
Comprehension	Picture	.492	.242	<.001*	.494	.002	—	.519	.025	.19
	Rhyme				.517	.025	.19	.522	.005	—
	Verify				.505	.013	—	.530	.026	.18

Note. Picture = Find the Picture; Rhyme = Find the Rhyme; Sig. = Significant.
* $p < .05$. *p* Values greater than .2 are not reported to facilitate the table's readability.

step, masked performance was entered in the second step and unmasked in the third. We then computed the unique variance associated with unmasked performance by subtraction.

This approach also addresses a concern with the previous analysis, that adding performance on masked versions may have simply increased the reliability of our experimental tasks because more trials contributed to the analysis. If this were the case, unmasked and masked versions should uniquely explain a similar proportion of the outcome variance. In contrast, if the pattern of significance differed for masked and unmasked tasks, this would suggest that masked versions tap different factors than the unmasked ones.

Results are shown in Table 4 and Figure 4. For oral fluency, only masked performance on Find the Picture and Find the Rhyme contributed significant unique variance (Figure 4A); there was no unique variance associated with unmasked variants. Although shared variance of masked and unmasked performance was low for Find the Picture ($\Delta R^2 = 0.018$), it was high for Find the Rhyme ($\Delta R^2 = 0.096$).

A similar story emerged for silent fluency. Only unique variance of the masked variant of Find the Picture reached significance (Figure 4B). Unmasked versions predicted no unique variance. As before, shared variance was particularly pronounced in Find the Rhyme ($\Delta R^2 = 0.190$) in contrast to the lower shared variance for the Find the Picture and the Verify tasks. This was also true for the two other reading measures, suggesting the Find the Rhyme task may have demanded similar skills independently of masking.

For decoding, only unmasked performance in Find the Rhyme and Verify accounted for significant amounts of unique variance (Figure 4C). This finding is intriguing, as masked performance did not contribute any unique variance to this task, suggesting that decoding may largely depend on acquired knowledge, not its automaticity. Finally, neither unique variance of unmasked nor masked performance accounted for variation in reading comprehension (Figure 4D).

Overall, this analysis suggests that results from the hierarchical regression are not simply due to the addition of trials. Importantly, masked performance captures a unique component of variance in fluency; whereas, unmasked performance captures unique variance in decoding. This apparent dissociation indicates that masked and unmasked performance may tap different components of skilled reading.

Question 2: Automaticity in the O → P Pathway

As described, it is difficult to isolate automaticity to any given pathway because both the O → P → S and O → S paths can be used to read a word. However, automaticity in the O → P pathway can be isolated using nonwords as there is no semantic route for recognizing them. Thus, to assess automaticity along this pathway, we split the data for Verify and Find the Rhyme by item-type and repeated the same analysis for nonwords only (Table 5; see Supplement S2 in the online supplemental material for equivalent analysis for words).

As before, masked performance on Find the Rhyme was significantly associated with oral fluency ($\Delta R^2 = 0.151$). This analysis did not reach significance in a parallel words-only analysis (see Supplement S2 in the online supplemental material), suggesting that masked Find the Rhyme may tap automaticity in the O → P pathway. For silent fluency, there was no unique variance associated with the masked performance of either task. This may be because phonological processing, which is isolated by the nonword trials, may be not as strongly engaged by measures of silent reading.

When the analysis of the decoding measures only included nonwords trials, results were the same as for analyses with all trials, with no strong effect of automaticity. Although all unmasked versions of tasks accounted for significant variance, none of the masked versions predicted decoding. This is particularly noteworthy for the Find the Rhyme nonword trials ($\Delta R^2 = 0.003$). Find the Rhyme itself was intended to maximize reliance on the O → P pathway, and the restriction to nonwords trials would increase its reliance on phonology or decoding skills. Thus, decoding, at least as measured by the WRMT, does not seem to be as strongly related to automaticity. Finally, reading comprehension was not predicted by unmasked or masked versions, underscoring that the initial finding was not driven by nonwords (see Table 5).

Table 4
Overview of Communality Analysis Including Shared and Unique Variance for Masked and Unmasked Tasks

		Overall	Shared	Unmasked		Masked	
Outcome	Task	R^2	ΔR^2_Δ	ΔR^2	Sig. ΔF	R^2	Sig. ΔF
Spoken Fluency	Picture	.137	.018	.006	—	.08	.032*
	Rhyme	.207	.096	.004	—	.073	.033*
	Verify	.158	.070	.026	—	.029	.189
Silent Fluency	Picture	.139	.028	.025	—	.069	.046*
	Rhyme	.241	.190	.023	—	.012	—
	Verify	.154	.078	.031	.173	.029	.188
Decoding	Picture	.243	.008	.058	.051	.001	—
	Rhyme	.434	.176	.082	.008*	0	—
	Verify	.401	.091	.132	.001*	.002	—
Comprehension	Picture	.42	.002	.000	—	.025	.186
	Rhyme	.272	.024	.001	—	.005	—
	Verify	.281	.013	.000	—	.026	.176

Note. Picture = Find the Picture; Rhyme = Find the Rhyme; Sig.= Significant.
* $p < .05$. p Values greater than .2 are not reported to facilitate the table's readability.

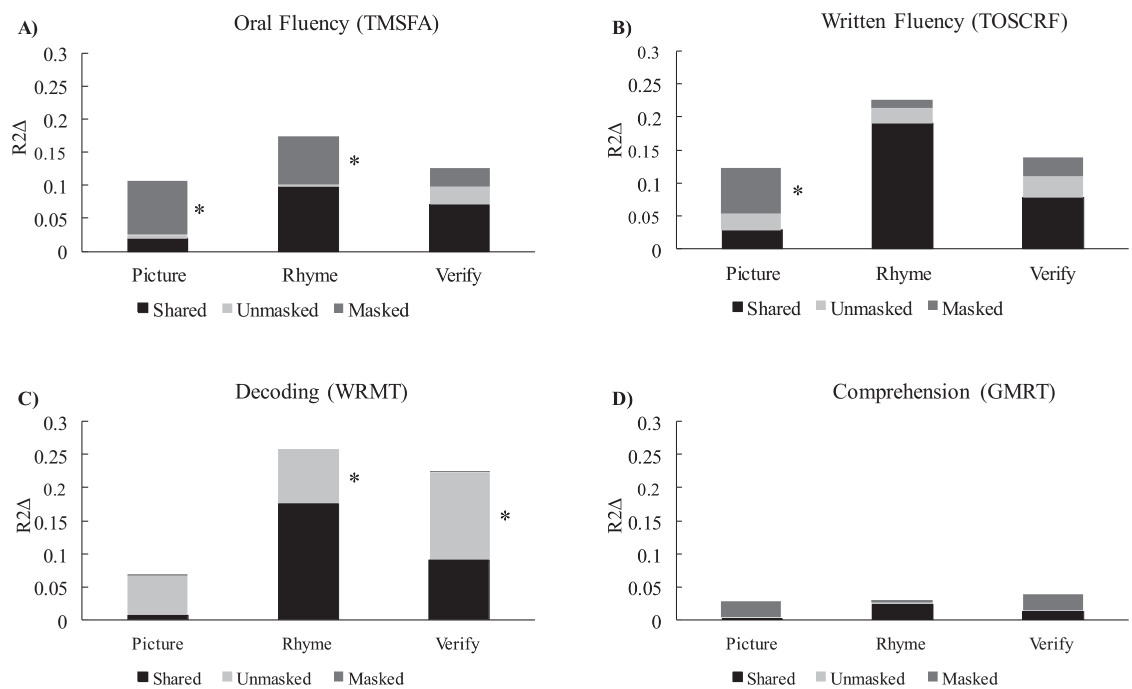

Figure 4. Unique and shared variance associated with masked and unmasked performance for all three tasks and reading outcomes. The asterisk (*) indicates unique variance reaching significance. Note that the PPVT is not included in this figure; thus, the changes in R^2 do not represent the total variance that is accounted. (A) Oral fluency (TMSFA = Texas Middle School Fluency Assessment); (B) Written fluency (TOSCRF = Test of Silent Contextual Reading Fluency); (C) Decoding (WRMT = Woodcock Johnson Reading Mastery Test); (D) Comprehension (GMRT = Gates-MacGinitie Reading Test); PPVT = Peabody Picture Vocabulary Test.

Discussion

Questions 1 and 2: Predicting Outcomes From Automaticity of Word Recognition

Our hierarchical regressions asked whether performance under masked presentation accounted for additional variance over and above unmasked performance for each of our outcome measures. This was observed for measures of word fluency when predicted by Find the Picture and Find the Rhyme (though not for Verify).

Automaticity in Find the Picture predicted both oral and silent fluency. As pointed out previously, Find the Picture may rely more on direct activation of semantics from orthography (the O → S pathway), and the other tasks may rely more on indirect activation flow via phonology (the O → P → S pathway). This suggests that reliance on the O → S pathway may be particularly important for automaticity and thus fluency. Moreover, these results are consistent with the notion that fluency in the reading system is achieved by direct activation of semantic information from orthography (e.g., Harm et al., 2004), rather than the slower decoding route.

Table 5
Overview of Regression Analyses for Nonwords (NW) Only for Find the Rhyme

			1. Vocabulary			2. Unmasked			3. Masked		
Outcome	Task	Item	R^2	ΔR^2_Δ	Sig. ΔF	R^2	ΔR^2_Δ	Sig. ΔF	R^2	ΔR^2_Δ	Sig. ΔF
Spoken Fluency	Rhyme	NW	.182	.033	.18	.347	.087	.026*	.521	.151	.002*
	Verify	NW				.348	.088	.025*	.379	.022	—
Silent Fluency	Rhyme	NW	.127	.016	—	.472	.207	<.001*	.503	.03	.16
	Verify	NW				.373	.123	.008*	.374	.001	—
Decoding	Rhyme	NW	.42	.176	.001*	.65	.246	<.001*	.652	.003	—
	Verify	NW				.641	.235	<.001*	.641	<.001	—
Comprehension	Rhyme	NW	.492	.242	<.001*	.511	.02	—	.537	.027	.17
	Verify	NW				.507	.015	—	.524	.018	—

Note. Rhyme = Find the Rhyme and Verify; Sig. = Significant.
* $p < .05$. p Values greater than .2 are not reported to facilitate the table's readability.

A subtly different relationship was observed for Find the Rhyme. Whereas masked Find the Rhyme showed a significant relationship with oral fluency, it did not predict silent fluency. When we replicated this analysis for nonwords (which isolate the O → P pathway), it became clear that this relationship was carried by nonword performance. This suggests that—even though reliance on the O → S pathway may be a more robust predictor of how fluent reading is—children can vary in the automaticity of the O → P pathway, and this matters for oral fluency. In contrast, silent fluency (which does not require reading aloud) may simply not require as much contribution from phonology. Importantly, neither oral nor silent fluency were uniquely predicted by unmasked performance in any task as indicated by the communality analysis.

A different pattern was observed for decoding (WRMT). Even restricting the analysis to nonwords only or entering masked performance first in the communality analysis (subsuming shared variance with unmasked), masked performance did not predict decoding. In contrast, the communality analysis indicated that unmasked versions of Find the Rhyme and Verify contributed uniquely to decoding scores, and Find the Picture made a marginal contribution. Unmasked task versions may better capture decoding scores because WRMT largely is an offline measure of decoding knowledge, not how quickly or automatically words are read. That is, it does not matter how automatic a child is for performing well in these measures, it only matters whether they know the words and relevant GPC regularities.

Interestingly, masked performance did not reliably account for differences in comprehension. This is surprising given the theoretical importance of automaticity to reading comprehension (e.g., LaBerge & Samuels, 1974; Perfetti, 1985). Because the GMRT involves students silently reading passages at their own pace, it may not require a speeded response and, therefore, not rely as heavily on the automaticity of single word recognition. Rather, most of the variance may have been accounted for by differences in oral language captured by the PPVT, which was highly correlated with the GMRT ($R = 0.49$, see Supplement S1 in the online supplemental material). However, even the unmasked versions of each task failed to capture much variance in comprehension, potentially reflecting that comprehension is not simply a product of efficient and accurate word recognition, but rather a variety of skills. This is consistent with existing reading research and the Simple View of Reading, suggesting that the most important factors in predicting reading comprehension are word recognition and listening comprehension (e.g., Hoover & Gough, 1990). However, it also may point to the need for better assessments, as Cutting and Scarborough (2006) have suggested. Although there are excellent standardized measures of children's ability to map print to sound for isolated words (e.g., the WRMT), there are few measures that tap children's ability to map print to meaning for isolated words.

Overall, automaticity showed a specific predictive relationship with outcomes. Automaticity was uniquely related to fluency. However, it did not account for unique variance in other reading skills. In contrast, decoding was best predicted by unmasked performance (with little unique variance associated with masked). The seeming double dissociation observed here—in which masked performance predicts fluency but not decoding and unmasked predicts decoding but not fluency—suggests the possibility of two crucial constructs in reading: a child's knowledge of the relevant material, and the automaticity with which it can be accessed (Stanovich, 1980). The dissociation also implies a form of discriminant validity to our measure of automaticity. Consistent with theories of automaticity in reading (LaBerge & Samuels, 1974; Perfetti, 1985), masked performance was related to outcomes that have been hypothesized to be reliant on sufficient automaticity, but not outcomes that do not have strong theoretical reasons to be tied to automaticity.

Moreover, we observed particularly strong relationships between automaticity on the direct pathway from orthography to semantics (assessed by Find the Picture) and both oral and silent fluency. Reliance on this pathway may be a critical factor for fluency but not decoding. However, this is not to say that O → S mappings are the only route to fluency. The significant effect for the nonwords in Find the Rhyme offers clear evidence that variation in the automaticity of the O → P route also may be important for fluency more generally. Thus, this pattern of results is consistent with the idea that automaticity can exist in different subprocesses of the reading system.

Further work should build on these results with finer-grained analyses of fluency, including an examination of the relationship between our masking-based measure of automatic word recognition and measures like RAN. Moreover, developmental work could ask if gains in fluency associated with automatic word recognition are linked to later gains in comprehension (LaBerge & Samuels, 1974; Perfetti, 1985; Schwanenflugel et al., 2006). Our results also may be meaningful for a larger population of younger students who still are developing automaticity in word reading and decoding. In fact, the predictive relationship of differences in automaticity to reading outcomes could be stronger in a population of children with a wider range of abilities. Finally, future work should explore how the role of automaticity may differ in populations that are diagnosed with reading difficulties. Prior research indicates that students with reading disabilities are often indistinguishable from those who are low achieving in reading but not identified with a learning disability (Fuchs, Fuchs, Mathes, Lipsey, & Roberts, 2001; O'Malley, Francis, Foorman, Fletcher, & Swank, 2002), suggesting that a similar pattern of results may be found in students with reading impairments. Nevertheless, as our sample includes a subset of students who receive additional educational services (e.g., an IEP) for unidentified reasons, our data cannot speak to how important automaticity is in predicting reading outcomes in individuals with learning disabilities (such as a language impairment or dyslexia).

Implications for Educational Practice

Although the present study is exploratory, it has implications for understanding and assessing reading skills. It supports the use of backward masking as a potential tool for probing automaticity of written word recognition and as a complement to established measures like RAN or speed of processing. Automaticity is clearly an important component of fluency (or perhaps precursor to it). Therefore, assessments targeting this process may yield insight into the challenges faced by struggling or emerging readers. Our analyses of these tasks suggest that there may be value in considering automaticity not as a single property of written word recognition, but as something that can be assessed in multiple pathways.

Assessments taking this approach may improve the ability to target intervention to the needed pathway.

In practice, it can be difficult to differentiate a struggling reader who has the relevant decoding knowledge but is not yet automatic from one who lacks the knowledge. This is because of the typically high correlation between standard measures of fluency and decoding (as we observed here, see the Supplement S1 in the online supplemental material). However, unmasked performance on our tasks contributed almost no variance to fluency, and masked performance contributed almost no variance to decoding. These findings suggest that these two variants together could capture a more orthogonal dimensionalization of a child's ability. That is, we may be able to use unmasked performance as a proxy for knowledge and use masked performance to assess automaticity. This approach has promise for identifying students that specifically struggle with automaticity. In addition, our findings suggest that conceptualizing knowledge and automaticity as two different constructs help understand a student's skill profile.

Limitations

The most important limitation of this study is that it is correlational. Thus, it is not possible to draw strong conclusions as to the specificity and causality of the relationships between automaticity and fluency. However, given that it may be hard to investigate these findings experimentally, this is a good starting point to explore how automaticity may differ, for instance, across different reading pathways.

As discussed in the introduction, there are clear reasons to use backward masking to test automaticity. However, backward masking may be particularly sensitive to lapses in attention. If one does not attend to the target word in the short interval it is presented, it is impossible to recover the information. Here, this may have been exacerbated by two factors. First, our sample of below-average middle schoolers may be more likely to show lower levels of executive function (Cutting, Materek, Cole, Levine, & Mahone, 2009). Second, the lack of fixation cues (e.g., a cross or button at the beginning of a trial) before masked trials may have made it more likely for students to miss a target word. At this point, we cannot dismiss the possibility that lapses in attention impacted performance. However, each iteration of a task version only included 20 trials and data collection was distributed over several days, allowing students to take breaks regularly. This should have minimized such lapses. Moreover, attentional lapses cannot explain why the masking effect was more pronounced for nonwords than words (because a lapse of attention would precede the target stimulus). Similarly, it does not predict correlations with fluency but not decoding—no evidence suggests vigilance is related to one but not the other. Similarly, we found correlations with Find the Picture, but not Verify when attentional lapses should have affected both equally. This pattern—a form of discriminant validity—is too specific to be the result of broad-based factors such as attention lapses.

One additional limitation of this study is that the stimuli were exclusively monosyllabic words. Monosyllabic words allowed better control over stimulus characteristics (e.g., GPC classes), and allowed us to choose competitors that were more closely matched (multisyllabic words are less likely to have close orthographic competitors). This was particularly important in Find the Rhyme, where competitors came from several classes (e.g., onset, offset). Nevertheless, multisyllabic words are a natural and frequent aspect of reading in middle-school and later. If they had been included in this study, accuracy would likely have been lower, and we suspect backward masking would have had larger effects, potentially revealing an even greater role for automaticity. Thus, our use of monosyllabic words may have resulted in a more conservative estimate of the role of automaticity in fluency. Given that multisyllabic words may be more appropriate for this age group and with readers of average or above-average ability (which were not tested here), future studies should explore whether automaticity in recognizing multisyllabic words follows similar mechanics as in monosyllables.

Conclusions

This study found that performance on masked task versions, likely tapping automaticity, predicted reading fluency in below-average middle-school students. This link was observed most strongly in a task that maximized reliance on the direct pathway from orthography to meaning. However, a similar finding with nonwords in a more phonological task suggested that orthographic-to-phonological pathways may also achieve automaticity and contribute to fluency. Automaticity did not predict variation in reading comprehension or decoding, suggesting a fairly precise link between automatic written word recognition and fluency, even as knowledge (measured by masked task versions) showed a strong relationship with decoding. Overall, these results suggest that conceptualizing automaticity as a nonbinary concept that may exist in more than one reading pathway may be helpful in developing better assessments, offering a new road for improving fluency, and perhaps reading in general.

References

Boersma, P., & Weenink, D. (2016). *Praat: Doing phonetics by computer*. Retrieved from http://www.praat.org/

Breitmeyer, B. G., & Ogmen, H. (2000). Recent models and findings in visual backward masking: A comparison, review, and update. *Perception & Psychophysics, 62,* 1572–1595. http://dx.doi.org/10.3758/BF03212157

Bryant, P. E., MacLean, M., Bradley, L. L., & Crossland, J. (1990). Rhyme and alliteration, phoneme detection, and learning to read. *Developmental Psychology, 26,* 429–438. http://dx.doi.org/10.1037/0012-1649.26.3.429

Cain, K., Oakhill, J. V., Barnes, M. A., & Bryant, P. E. (2001). Comprehension skill, inference-making ability, and their relation to knowledge. *Memory & Cognition, 29,* 850–859. http://dx.doi.org/10.3758/BF03196414

Carreiras, M., Perea, M., & Grainger, J. (1997). Effects of orthographic neighborhood in visual word recognition: Cross-task comparisons. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 23,* 857–871. http://dx.doi.org/10.1037/0278-7393.23.4.857

Catts, H. W., Fey, M. E., Zhang, X., & Tomblin, J. B. (1999). Language basis of reading and reading disabilities: Evidence from a longitudinal investigation. *Scientific Studies of Reading, 3,* 331–361. http://dx.doi.org/10.1207/s1532799xssr0304_2

Chace, K. H., Rayner, K., & Well, A. D. (2005). Eye movements and phonological parafoveal preview: Effects of reading skill. *Canadian Journal of Experimental Psychology / Revue Canadienne de Psychologie Experimentale, 59,* 209–217. http://dx.doi.org/10.1037/h0087476

Cirino, P. T., Romain, M. A., Barth, A. E., Tolar, T. D., Fletcher, J. M., & Vaughn, S. (2013). Reading skill components and impairments in middle school struggling readers. *Reading and Writing, 26,* 1059–1086. http://dx.doi.org/10.1007/s11145-012-9406-3

Coltheart, M., Rastle, K., Perry, C., Langdon, R., & Ziegler, J. (2001). DRC: A dual route cascaded model of visual word recognition and reading aloud. *Psychological Review, 108,* 204–256. http://www.ncbi.nlm.nih.gov/pubmed/11212628. http://dx.doi.org/10.1037/0033-295X.108.1.204

Compton, D. L. (2003). The influence of item composition on RAN letter performance in first-grade children. *The Journal of Special Education, 37,* 81–94. http://dx.doi.org/10.1177/00224669030370020301

Curtis, M. E. (1980). Development of components of reading skill. *Journal of Educational Psychology, 72,* 656–669. http://dx.doi.org/10.1037/0022-0663.72.5.656

Cutting, L. E., Materek, A., Cole, C. A. S., Levine, T. M., & Mahone, E. M. (2009). Effects of fluency, oral language, and executive function on reading comprehension performance. *Annals of Dyslexia, 59,* 34–54. http://dx.doi.org/10.1007/s11881-009-0022-0

Cutting, L. E., & Scarborough, H. S. (2006). Prediction of reading comprehension: Relative contributions of word recognition, language proficiency, and other cognitive skills can depend on how comprehension is measured. *Scientific Studies of Reading, 10,* 277–299. http://dx.doi.org/10.1207/s1532799xssr1003_5

Davis, C. J., & Lupker, S. J. (2006). Masked inhibitory priming in english: Evidence for lexical inhibition. *Journal of Experimental Psychology: Human Perception and Performance, 32,* 668–687. http://dx.doi.org/10.1037/0096-1523.32.3.668

Denckla, M. B., & Rudel, R. (1974). Rapid "automatized" naming of pictured objects, colors, letters and numbers by normal children. *Cortex, 10,* 186–202. http://dx.doi.org/10.1016/S0010-9452(74)80009-2

Denckla, M. B., & Rudel, R. G. (1976). Rapid "automatized" naming (R. A. N): Dyslexia differentiated from other learning disabilities. *Neuropsychologia, 14,* 471–479. http://dx.doi.org/10.1016/0028-3932(76)90075-0

Duff, F. J., Reen, G., Plunkett, K., & Nation, K. (2015). Do infant vocabulary skills predict school-age language and literacy outcomes? *Journal of Child Psychology and Psychiatry, 56,* 848–856. http://dx.doi.org/10.1111/jcpp.12378

Dunn, L. M., & Dunn, D. M. (2007). *Peabody Picture Vocabulary Test,* 4th ed., *Summary* (Vol. 30). San Antonio, TX: Pearson Education, Inc. http://dx.doi.org/10.1037/t15144-000

Ehri, L. C. (1995). Phases of development in learning to read words by sight. *Journal of Research in Reading, 18,* 116–125. http://dx.doi.org/10.1111/j.1467-9817.1995.tb00077.x

Ehri, L. C., & Wilce, L. S. (1983). Development of word identification speed in skilled and less skilled beginning readers. *Journal of Education & Psychology, 75,* 3–18. http://dx.doi.org/10.1037/0022-0663.75.1.3

Förster, J., Higgins, E. T., & Bianco, A. T. (2003). Speed/accuracy decisions in task performance: Built-in trade-off or separate strategic concerns? *Organizational Behavior and Human Decision Processes, 90,* 148–164. http://dx.doi.org/10.1016/S0749-5978(02)00509-5

Forster, K. I., & Davis, C. (1984). Repetition priming and frequency attenuation in lexical access. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 10,* 680–698. http://dx.doi.org/10.1037/0278-7393.10.4.680

Francis, D. J., Barth, A., Cirino, P., Reed, D. K., & Fletcher, J. M. (2010). *Texas Middle School Fluency Assessment* (Version 2.0). Houston, TX: University of Houston/Texas Education Agency.

Frost, R., & Yogev, O. (2001). Orthographic and phonological computation in visual word recognition: Evidence from backward masking in Hebrew. *Psychonomic Bulletin & Review, 8,* 524–530. http://dx.doi.org/10.3758/BF03196188

Fuchs, D., Fuchs, L. S., Mathes, P. G., Lipsey, M. W., & Roberts, P. H. (2001, August). *Is "learning disabilities" just a fancy term for low achievement? A meta-analysis of reading differences between low achievers with and without the label: Executive Summary*. Paper presented at the Learning Disabilities Summit: Building a Foundation for the Future, Washington, DC.

Georgiou, G. K., Parrila, R., Cui, Y., & Papadopoulos, T. C. (2013). Why is rapid automatized naming related to reading? *Journal of Experimental Child Psychology, 115,* 218–225. http://dx.doi.org/10.1016/j.jecp.2012.10.015

Goswami, U., Gombert, J. E., & de Barrera, L. F. (1998). Children's orthographic representations and linguistic transparency: Nonsense word reading in English, French, and Spanish. *Applied Psycholinguistics, 19,* 19–52. http://dx.doi.org/10.1017/S0142716400010560

Gough, P. B., & Tunmer, W. E. (1986). Decoding, reading, and reading disability. *Remedial and Special Education, 7,* 6–10. http://dx.doi.org/10.1177/074193258600700104

Hammill, D. D., Wiederholt, J. L., & Allen, E. A. (2006). *TOSCRF: Test of Silent Contextual Reading Fluency: Examiner's manual*. Austin, TX: Pro-Ed, Inc.

Hammill, D. D., Wiederholt, J. L., & Allen, E. A. (2014). *Test of Silent Contextual Reading Fluency* (2nd ed.). Austin, TX: Pro-Ed, Inc.

Harm, M. W., Seidenberg, M. S., Macdonald, M., Thornton, R., & Zevin, J. (2004). Computing the meanings of words in reading: Cooperative division of labor between visual and phonological processes. *Psychological Review, 111,* 662–720. http://dx.doi.org/10.1037/0033-295X.111.3.662

Hock, M. F., Brasseur, I. F., Deshler, D. D., Catts, H. W., Marquis, J. G., Mark, C. A., & Stribling, J. W. (2009). What is the reading component skill profile of adolescent struggling readers in urban schools. *Learning Disability Quarterly, 32,* 21–38. Retrieved from http://journals.sagepub.com/doi/pdf/10.2307/25474660

Hogaboam, T., & Perfetti, C. A. (1978). Reading skill and the role of verbal experience in decoding. *Journal of Educational Psychology, 70,* 717–729. http://dx.doi.org/10.1037/0022-0663.70.5.717

Hoover, W. A., & Gough, P. B. (1990). The simple view of reading. *Reading and Writing, 2,* 127–160. http://dx.doi.org/10.1007/BF00401799

Jenkins, J. R., Fuchs, L. S., van den Broek, P., Espin, C., & Deno, S. L. (2003). Sources of individual differences in reading comprehension and reading fluency. *Journal of Educational Psychology, 95,* 719–729. http://dx.doi.org/10.1037/0022-0663.95.4.719

Joshi, M. R., & Aaron, P. G. (2000). The component model of reading: Simple view of reading made a little more complex. *Reading Psychology, 21,* 85–97. http://dx.doi.org/10.1080/02702710050084428

Juel, C. (1983). The development and use of mediated word identification. *Reading Research Quarterly, 18,* 306. http://dx.doi.org/10.2307/747391

Kail, R. (2000). Speed of information processing: Developmental change and links to intelligence. *Journal of School Psychology, 38,* 51–61. http://dx.doi.org/10.1016/S0022-4405(99)00036-9

Kail, R., & Hall, L. K. (1994). Processing speed, naming speed, and reading. *Developmental Psychology, 30,* 949–954. http://dx.doi.org/10.1037/0012-1649.30.6.949

Kim, Y. G. (2015). Developmental, component-based model of reading fluency: An investigation of predictors of word-reading fluency, text-reading fluency, and reading comprehension. *Reading Research Quarterly, 50,* 459–481. http://dx.doi.org/10.1002/rrq.107

Kirby, J. R., Georgiou, G. K., Martinussen, R., Parrila, R., Bowers, P., & Landerl, K. (2010). Naming speed and reading: From prediction to instruction. *Reading Research Quarterly, 45,* 341–362. http://dx.doi.org/10.1598/RRQ.45.3.4

Klauda, S. L., & Guthrie, J. T. (2008). Relationships of three components of reading fluency to reading comprehension. *Journal of Educational*

Psychology, 100, 310–321. http://dx.doi.org/10.1037/0022-0663.100.2.310

Kuhn, M. R., Schwanenflugel, P. J., Meisinger, E. B., Levy, B. A., & Rasinski, T. V. (2010). Aligning theory and assessment of reading fluency: Automaticity, prosody, and definitions of fluency. *Reading Research Quarterly, 45,* 230–251. http://dx.doi.org/10.1598/RRQ.45.2.4

LaBerge, D., & Samuels, S. J. (1974). Toward a theory of automatic information processing in reading. *Cognitive Psychology, 6,* 293–323. http://dx.doi.org/10.1016/0010-0285(74)90015-2

Leinenger, M. (2014). Phonological coding during reading. *Psychological Bulletin, 140,* 1534–1555. http://dx.doi.org/10.1037/a0037830

Lervåg, A., & Hulme, C. (2009). Rapid automatized naming (RAN) taps a mechanism that places constraints on the development of early reading fluency. *Psychological Science, 20,* 1040–1048. http://dx.doi.org/10.1111/j.1467-9280.2009.02405.x

Logan, G. (1997). Automaticity and reading: Perspectives from the instance theory of automatization. *Reading & Writing Quarterly: Overcoming Learning Difficulties, 13,* 123–146. http://dx.doi.org/10.1080/1057356970130203

Luce, R. D. (1986). *Response times: Their role in inferring elementary mental organization*. New York, NY: Oxford University Press on Demand.

Luck, S. J., & Vogel, E. K. (1997). The capacity of visual working memory for features and conjunctions. *Nature, 390,* 279–281. http://dx.doi.org/10.1038/36846

MacGinitie, W. H., MacGinitie, R. K., Maria, K., & Dreyer, L. (2000). *Gates-MacGinitie Reading Tests* (4th ed.). Chicago, IL: Riverside Publishing.

McCardle, P., Scarborough, H. S., & Catts, H. W. (2001). Predicting, explaining, and preventing children's reading difficulties. *Learning Disabilities Research & Practice, 16,* 230–239. http://dx.doi.org/10.1111/0938-8982.00023

McMurray, B., Samelson, V. M., Lee, S. H., & Tomblin, J. B. (2010). Individual differences in online spoken word recognition: Implications for SLI. *Cognitive Psychology, 60,* 1–39. http://dx.doi.org/10.1016/j.cogpsych.2009.06.003

Miller, C. A., Kail, R., Leonard, L. B., & Tomblin, J. B. (2001). Speed of processing in children with specific language impairment. *Journal of Speech, Language, and Hearing Research, 44,* 416–433. http://dx.doi.org/10.1044/1092-4388(2001/034)

Moll, K., Fussenegger, B., Willburger, E., & Landerl, K. (2009). RAN is not a measure of orthographic processing. Evidence from the asymmetric German orthography. *Scientific Studies of Reading, 13,* 1–25. http://dx.doi.org/10.1080/10888430802631684

Nag, S., & Snowling, M. J. (2012). Reading in an alphasyllabary: Implications for a language universal theory of learning to read. *Scientific Studies of Reading, 16,* 404–423. http://dx.doi.org/10.1080/10888438.2011.576352

Nation, K., & Snowling, M. J. (1998). Semantic processing and the development of word-recognition skills: Evidence from children with reading comprehension difficulties. *Journal of Memory and Language, 39,* 85–101. http://dx.doi.org/10.1006/jmla.1998.2564

Nation, K., & Snowling, M. J. (2000). Factors influencing syntactic awareness skills in normal readers and poor comprehenders. *Applied Psycholinguistics, 21,* 229–241. http://dx.doi.org/10.1017/S0142716400002046

National Governors Association Center for Best Practices & Council of Chief State School Officers. (2010). *Common Core state standards*. Washington, DC: Author.

National Reading Panel. (2000). *Teaching children to read: An evidence-based assessment of the scientific research literature on reading and its implications for reading instruction*. Washington, DC: National Institute of Child Health and Human Development.

Nunes, T., Bryant, P., & Barros, R. (2012). The development of word recognition and its significance for comprehension and fluency. *Journal of Educational Psychology, 104,* 959–973. http://dx.doi.org/10.1037/a0027412

O'Malley, K. J., Francis, D. J., Foorman, B. R., Fletcher, J. M., & Swank, P. R. (2002). Growth in precursor and reading-related skills: Do low-achieving and IQ-discrepant readers develop differently? *Learning Disabilities Research & Practice, 17,* 19–34. http://dx.doi.org/10.1111/1540-5826.00029

Oslund, E. L., Clemens, N. H., Simmons, D. C., Smith, S. L., & Simmons, L. E. (2016). How vocabulary knowledge of middle-school students from low socioeconomic backgrounds influences comprehension processes and outcomes. *Learning and Individual Differences, 45,* 159–165. http://dx.doi.org/10.1016/j.lindif.2015.11.013

Pan, J., McBride-Chang, C., Shu, H., Liu, H., Zhang, Y., & Li, H. (2011). What is in the naming? A 5-year longitudinal study of early rapid naming and phonological sensitivity in relation to subsequent reading skills in both native Chinese and English as a second language. *Journal of Educational Psychology, 103,* 897–908. http://dx.doi.org/10.1037/a0024344

Perfetti, C. A. (1985). *Reading ability*. New York, NY: Oxford University Press.

Perfetti, C. A., & Bell, L. (1991). Phonemic activation during the first 40 ms of word identification: Evidence from backward masking and priming. *Journal of Memory and Language, 30,* 473–485. http://dx.doi.org/10.1016/0749-596X(91)90017-E

Perfetti, C. A., Bell, L. C., & Delaney, S. M. (1988). Automatic (prelexical) phonetic activation in silent word reading: Evidence from backward masking. *Journal of Memory and Language, 27,* 59–70. http://dx.doi.org/10.1016/0749-596X(88)90048-4

Perfetti, C. A., & Hogaboam, T. (1975). Relationship between single word decoding and reading comprehension skill. *Journal of Educational Psychology, 67,* 461–469. http://dx.doi.org/10.1037/h0077013

Plaut, D. C., McClelland, J. L., & Seidenberg, M. S. (1995). Reading exception words and pseudowords: Are two routes really necessary. *Connectionist Models of Memory and Language,* 145–159. http://dx.doi.org/10.4324/9781315794495

Powell, D., Stainthorp, R., Stuart, M., Garwood, H., & Quinlan, P. (2007). An experimental comparison between rival theories of rapid automatized naming performance and its relationship to reading. *Journal of Experimental Child Psychology, 98,* 46–68. http://dx.doi.org/10.1016/j.jecp.2007.04.003

Rasinski, T. V., Padak, N. D., McKeon, C. A., Wilfong, L. G., Friedauer, J. A., & Heim, P. (2005). Is reading fluency a key for successful high school reading? *Journal of Adolescent & Adult Literacy, 49,* 22–27. http://dx.doi.org/10.1598/JAAL.49.1.3

Reicher, G. M. (1969). Perceptual recognition as a function of meaninfulness of stimulus material. *Journal of Experimental Psychology, 81,* 275–280. http://dx.doi.org/10.1037/h0027768

Roberts, M. A., Rastle, K., Coltheart, M., & Besner, D. (2003). When parallel processing in visual word recognition is not enough: New evidence from naming. *Psychonomic Bulletin & Review, 10,* 405–414. http://dx.doi.org/10.3758/BF03196499

Scarborough, H. S. (1990). Very early language deficits in dyslexic children. *Child Development, 61,* 1728–1743. http://dx.doi.org/10.2307/1130834

Schwanenflugel, P. J., Meisinger, E. B., Wisenbaker, J. M., Kuhn, M. R., Strauss, G. P., & Morris, R. D. (2006). Becoming a fluent and automatic reader in the early elementary school years. *Reading Research Quarterly, 41,* 496–522. http://dx.doi.org/10.1598/RRQ.41.4.4

Seidenberg, M. S. (2005). Connectionist models of word reading. *Current Directions in Psychological Science, 14,* 238–242. http://dx.doi.org/10.1111/j.0963-7214.2005.00372.x

Seidenberg, M. S., & McClelland, J. L. (1989). A distributed, developmental model of word recognition and naming. *Psychological Review, 96,* 523–568. http://dx.doi.org/10.1037/0033-295X.96.4.523

Share, D. L. (1995). Phonological recoding and self-teaching: Sine qua non of reading acquisition. *Cognition, 55,* 151–218. http://dx.doi.org/10.1016/0010-0277(94)00645-2

Share, D. L. (2008). Orthographic learning, phonological recoding, and self-teaching. *Advances in Child Development and Behavior, 36,* 31–82. http://dx.doi.org/10.1016/S0065-2407(08)00002-5

Stanovich, K. E. (1980). Toward an interactive-compensatory model of individual differences in the development of reading fluency. *Reading Research Quarterly, 16,* 32–71. http://dx.doi.org/10.2307/747348

Tamura, N., Castles, A., & Nation, K. (2017). Orthographic learning, fast and slow: Lexical competition effects reveal the time course of word learning in developing readers. *Cognition, 163,* 93–102. http://dx.doi.org/10.1016/j.cognition.2017.03.002

Texas Education Agency, University of Houston, & The University of Texas System. (2008). *Texas Middle School Fluency Assessment*. Austin, TX: Author.

Turvey, M. T. (1973). On peripheral and central processes in vision: Inferences from an information-processing analysis of masking with patterned stimuli. *Psychological Review, 80,* 1–52. http://dx.doi.org/10.1037/h0033872

Urry, K., Burns, N. R., & Baetu, I. (2015). Accuracy-based measures provide a better measure of sequence learning than reaction time-based measures. *Frontiers in Psychology, 6,* 1158. http://dx.doi.org/10.3389/fpsyg.2015.01158

Walczyk, J. J. (2000). The interplay between automatic and control processes in reading. *Reading Research Quarterly, 35,* 554–566. http://dx.doi.org/10.1598/RRQ.35.4.7

Wegener, S., Wang, H.-C., de Lissa, P., Robidoux, S., Nation, K., & Castles, A. (2017). Children reading spoken words: Interactions between vocabulary and orthographic expectancy. *Developmental Science,* e12577. http://dx.doi.org/10.1111/desc.12577

Wheeler, D. D. (1970). Processes in word recognition. *Cognitive Psychology, 1,* 59–85. http://dx.doi.org/10.1016/0010-0285(70)90005-8

Wijayathilake, M. A. D. K., & Parrila, R. (2013). Predictors of word reading skills in good and struggling readers in Sinhala. *Writing Systems Research, 6,* 120–131. http://dx.doi.org/10.1080/17586801.2013.846844

Woodcock, R. W., McGrew, K., & Mather, N. (2001). *Woodcock-Johnson Tests of Achievement* (3rd ed.). Itasca, IL: Riverside.

Received February 15, 2017
Revision received February 16, 2018
Accepted February 22, 2018 ■