Published in final edited form as:

Read Writ. 2008 April 1; 21(3): 177–203. doi:10.1007/s11145-007-9061-2.

The influence of reading unit size on the development of Stroop interference in early word decoding

Paula J. Schwanenflugel, Robin D. Morris, Melanie R. Kuhn, Gregory P. Strauss, and Jennifer M. Sieczko

P. J. Schwanenflugel · J. M. Sieczko, Department of Educational Psychology and Instructional Technology, University of Georgia, 325 R Aderhold Hall, Athens, GA 30602, USA, email: pschwan@uga.edu

R. D. Morris, Georgia State University, Atlanta, GA, USA

M. R. Kuhn, Rutgers University, New Brunswick, NJ, USA

G. P. Strauss, University of Nevada-Las Vegas, Las Vegas, NV, USA

Abstract

The purpose of the experiments was to determine the automatic use of large or small word reading units in young readers in the absence of word decoding strategies. Picture-word Stroop interference was examined from four types of conflicting labels: (a) words containing both highly predictable grapheme–phoneme correspondence (GPC) units and highly consistent rime units (henceforth, Hi-GPC + Hi-Rime); (b) words with highly predictable GPC units and less consistent rime units (Hi-GPC + Lo-Rime); (c) words with low predictability GPC units and highly consistent rime units (Low GPC + High Rime); (d) nonwords that contained both highly predictable GPC and highly consistent rime units. Naming time for pictures containing these labels was compared against that for pictures with random letter strings or no labels. In Experiment 1, Stroop interference was examined in first, second, and third grade children to determine whether there was developmental change in the presence of rime or GPC interference. In Experiment 2, Stroop interference was examined as a function of relative reading skill in first grade children. In Experiment 3, Stroop interference in adults was compared to the use of rime or GPC pronunciation strategies for nonword reading. In all experiments, Stroop interference in picture naming was longer for pictures with highly predictable GPC unit labels than less predictable GPC unit labels. However, in Experiment 3, even though adults showed interference from predictable GPC units in the Stroop task, they always preferred rime pronunciation for ambiguous non-words in the nonword reading task. It is argued that the current experiments provide evidence for a flexible units model.

Keywords

Stroop Task; Automaticity; Reading units; Word decoding

The attainment of quick and accurate word reading is fundamental to children's growth as readers. Quick and accurate word reading is often referred to as automatic word recognition, although the obligatory aspect of word reading is sometimes included in this definition (Logan, 1997). Most developmental theories of reading place the development of automatic word

Correspondence to: Paula J. Schwanenflugel.

The results of this study were presented at the Cognitive Development Society meeting, November 2001, Virginia Beach, VA, and the American Educational Research Association meeting, April 2004, San Diego, California.

recognition as a key step in becoming a skilled reader (Adams, 1990; Chall, 1996; LaBerge & Samuels, 1974).

Experimentally, the obligatory aspect of automatic word processing has traditionally been measured through the use of Stroop task (Stroop, 1935) and its variants. The Stroop task directly taxes the obligatory component of automatic word recognition by requiring participants to respond to a primary stimulus while ignoring conflicting print. The interference caused by attempting to name the ink color of a word (or picture) with a conflicting color term (or conflicting label) compared against a control lacking this conflict has been termed the “Stroop Effect” after J.R. Stroop (1935) who discovered it.

Several studies have linked the development of Stroop interference to the onset of reading (Ehri & Wilce, 1979; Guttentag & Haith, 1978; Stanovich, Cunningham, & West, 1981). Stroop interference changes as a function of developmental and individual differences in the development of reading skill as well, although this effect is not as well-established. Ehri (1976) found that early, proficient readers showed larger interference for picture stimuli with conflicting labels compared to late, less proficient readers. Similarly, Schadler and Thissen (1981) found that Stroop interference from words increased until children’s reading skills reached a fourth grade level psychometrically, after which interference declined. Golinkoff and Rosinski (1976) did not find reading skill differences in Stroop interference in third and fifth graders, but in later elementary school, children become better able inhibit stimulus features that threaten to distract them (Pritchard & Neumann, 2004; Tipper, Bourque, Anderson, & Brehaut, 1989). Although the overall size of the Stroop effect may change, the effect itself does not disappear across the life span (Comalli, Wapner, & Werner, 1962).

Given that Stroop interference develops along with reading skill, what exactly becomes processed obligatorily as reading becomes automatic? Some early findings imply that, not only are words processed obligatorily, but sublexical units below the level of individual words may also be. The evidence for this point comes from the fact that pronounceable nonwords also cause Stroop interference. Interference from pronounceable nonwords may emerge with the development of skilled reading. Guttentag and Haith (1978) found that semantically-driven word interference precedes the development of Stroop interference attributable to sublexical units found in pronounceable nonwords.

There is considerable debate regarding which units below the level of the written word are used by young readers in automatic word reading. Few reading researchers dispute the relevance of spelling-sound relations in learning to read. Spelling-sound relations are the relationships that exist between the phonemes heard in pronounced speech and various spelling (or orthographic) patterns that map onto that speech. Past research has yielded mixed results regarding the nature and size of the sublexical units that children and adults use as they map the written word to speech in word decoding (Coltheart & Leahy, 1992; Goswami, 1993).

Some researchers hypothesize that children begin the process of learning to read words with the use of small sublexical units that map letters onto distinct phonemes, called grapheme-phoneme correspondence rules (or GPC; Coltheart & Leahy, 1992). A grapheme is a single letter (e.g., *t*-) or letter cluster (*sh*-), that gets mapped on to a single phoneme or sound. Supporters of the small-units view believe that GPC strategies aid readers in making more successful attempts at correct pronunciation or spelling when they come across unfamiliar words (Perry, Ziegler, & Coltheart, 2002). We use the term *small unit* synonymously with *GPC* throughout. According to this view, children initially read using GPC units, making use of larger spelling-sound mappings only later when they have built up a large sight word vocabulary. In support of this view, Colheart and Leahy (1992), Bowey and Underwood (1996), and Manis, Szeszulski, Howell, and Horn (1986) found young children were more

affected by small unit consistency in nonword pronunciation than rime unit consistency compared to older children and adults. Moreover, Christensen and Bowey (2005) find that children progress faster when instructed using small-unit phonics strategies than large unit strategies. This proposed developmental change as a function of developing reading skill is sometimes called simply the *small-to-large unit progression* or *small-units-first* view (Duncan, Seymour, & Hill, 2000).

An alternative view is that children make use of larger sublexical units that either engage word analogies or rime units for word decoding (Goswami, Ziegler, Dalton, & Schneider, 2003). Rime units consist of recurrent rimes in spoken syllables. For single syllable words, this view organizes word parts into *onset* (the initial consonant) and *rime* units consisting of the vowel and subsequent consonants. Readers can produce more consistent and accurate pronunciations when they recognize a familiar rime in an unfamiliar word and utilize it for pronunciation (Treiman, Mullennix, Bijeljac-Babic, & Richmond-Welty, 1995). Throughout, we use the term *large unit* synonymously with *rime unit*. According to this view, awareness of large unit relations emerges early (MacLean, Bryant, & Bradley, 1987; Treiman & Zukowski, 1991) and this large unit knowledge is then used as a decoding strategy in early reading (Goswami, 1993). In favor of this view, it has been shown that beginning readers are more able to make use of rime analogies as “clues” (e.g., *beak*) presented prior to unfamiliar to-be-read words (such as *peak*) compared to other words that shared similarities at the GPC level (*bean*). Further, children may respond more quickly to rime-level instruction than grapheme–phoneme instruction (Bruck & Treiman, 1992; Walton & Walton, 2002). Moreover, children’s tendency to use rime-unit pronunciations in nonword reading tasks may be largely predicted by the number of orthographic “neighbors” it has (Treiman, Goswami, & Bruck, 1990). This proposed developmental change as a function of reading skill has sometimes been called *large-to-small units progression* or the *large-units first* view (Duncan et al., 2000).

Finally, others argue that both GPC and rime units are both available to the reader, but that they are used strategically in decoding depending on the requirements of the experimental task (Brown & Deavers, 1999). In fact, Brown and Deavers (1999) have shown that merely shifting the proportion of words in a list to those following small unit rules (e.g., *dock*) will encourage ambiguous nonwords to be pronounced using small units, whereas doing the reverse (e.g., *chalk*) will encourage the same nonwords to be pronounced using large units. That is, some nonwords (such as *dalk*) that can be pronounced either using a large unit approach (pronounced like *chalk*) or a small unit one (as in *talc*). Presumably, the pronunciation of *dalk* would depend heavily on whether the rest of the list contained many words such as *chalk* which require a rime unit pronunciation or ones like *talc* which conform to a small unit one. Further, Brown and Deavers showed that the use of GPC rules or rime units seems to be driven by strategies that are quite task-dependent and used flexibly to meet task demands. One model of word naming contains both rime-level and GPC-level features that are used probabilistically (Stuart, Masterson, Dixon, & Quinlan, 1999).

If these sublexical units (either rime or GPC features) are available automatically, it would seem that pronounceable nonwords (which must contain such units) should be obligatorily processed also, causing Stroop interference (Guttentag & Haith, 1978). The literature is mixed, however, with regards to the relative size of the interference effects from words and pronounceable nonwords. When contrasted against semantically unrelated word labels (to which interference from pronounceable nonwords can be most directly compared), some studies find similar interference from pronounceable nonwords as words (Berent & Marom, 2005; Redding & Gerjet, 1977; Rosinski, 1977), others find greater interference from words than pronounceable nonwords (Bakan & Alperson, 1967; Guttentag & Haith, 1978; Monsell, Taylor, & Murphy, 2001; Rayner & Posnansky, 1978), and others find no interference from pronounceable nonwords at all (Marmurek, Proctor, & Javor, 2006). Moreover, it is unclear

whether such interference would emerge simultaneously with the development of reading skill or as a later outgrowth of it as implied by the developmental course suggested by Guttentag and Haith (1978).

The purpose of the current experiments is to ascertain the nature of sublexical units used in automatic word reading in the absence of strategies that encourage their use. In the Stroop task, the stated task goal is to ignore the label and to name the primary stimulus. Strategies initiated around this task, if any, are developed to inhibit the processing of the conflicting label. If particular types of units (large or small or both) are activated automatically, then it would be anticipated that words containing them would cause differential interference with picture naming. Use of the Stroop task ensures a direct comparison of unit-size that is relatively unconfounded by strategic influence, and allows the unit-size dilemma to be compared at the most basic level of obligatory processing.

In the current set of experiments, picture-word Stroop interference was examined from four types of conflicting labels: (a) words with highly predictable GPC and highly consistent rime units (henceforth, High-GPC/High-Rime); (b) words with highly predictable GPC units but with inconsistent rime units (High-GPC/Low-Rime); (c) words with low predictability GPC units but with highly consistent rime units (Low GPC + High Rime); (d) nonwords containing High GPC + High Rime units (i.e., *keck*). If the large-units-first hypothesis is correct, then one might expect greater Stroop interference in young readers when labels activate highly consistent rime units. That is, young readers should display larger Stroop interference for pictures with Low-GPC + High-Rime word, High-GPC + Hi Rime word, and nonword labels than for High GPC + Low Rime labels. If the small-units-first hypothesis is correct, then the opposite would be true. Labels containing highly predictable GPC units should cause more interference. That is, young readers should show larger Stroop interference for pictures having High GPC + High Rime word, High GPC + Low Rime word, and nonword labels compared to Low GPC + High Rime word labels. If both types of units are processed obligatorily, then similar interference should be found for all types of word and nonword labels. However, it may be that semantically-driven word interference develops prior to nonword interference, as suggested by Guttentag and Haith (1978). In that case, it might be expected that younger and less skilled readers would show greater interference for words than nonwords. Presumably, for all models, interference differences between labels should disappear for older readers as both unit types become automatized.

In the first experiment, Stroop interference for different types of conflicting labels was examined in first, second, and third grade children to determine whether there was change in the presence of rime or GPC interference with age. In the second experiment, Stroop interference was examined as a function of relative reading skill in first grade children to determine if rime or GPC interference was more characteristic of the early stages of reading. In the third experiment, Stroop interference from rime or GPC units was examined at the other end of the developmental spectrum in adults and Stroop interference was compared to nonword reading for consistency in the types of units involved as a function of task.

Experiment 1

Method

Participants—Participants were 118 first grade, 84 second grade, and 79 third grade students from an urban sector of northeast Georgia and suburban central New Jersey. Each child had received informed consent from his or her parent or guardian, and had assented to participate. Children were native English speakers. The children participating were part of a larger study of reading fluency, and received the present task along with other standardized and experimental measures of reading fluency.

Apparatus—The stimuli were presented by means of a Dell Inspiron (8000) laptop computer using E-Prime experiment software (Psychology Software Tools, Inc., Version 5.0). The pictures were projected onto a 15" laptop computer screen, with a viewing distance of approximately two feet. Naming onset latencies were measured by a voice-operated microphone (Audio-Technica Cardioid low impedance model # ATR 20C), which was connected to a serial response box (Psychology Software Tools, Inc., model # 200A).

Stimulus materials—The pictures were 66 digitized line drawings of various animals, shapes, and objects provided in Cycowicz, Friedman, Rothstein, and Snodgrass (1997). All pictures were exact duplicates of those used in Cycowicz et al., and were provided by those researchers with consent to be used in this study. The pictures were designed for children between the ages of 5 and 7, and had nameability rates in that study of 90% or higher. The information printed within the pictures was prepared by means of Adobe Photoshop 5.0. Each word was written in lowercase, Arial Black font regular, size 18 points, in black color, and saved in bitmap format. The projected pictures and words printed within the pictures appeared on the computer screen as black line drawings against a white background, and each stimulus appeared on the computer screen for five seconds or until a word was spoken into the microphone. There were 66 stimuli, with 60 experimental and six practice stimuli, one modeled after each condition. The words printed on the various pictures were three to five letters long, with letter string stimuli also ranging from three to five letters in length. Stimuli from the picture control condition were not modified from their original format, as provided by Cycowicz et al. We had originally intended for there to be 10 items per experimental condition, however, some reshuffling was necessary after the word "rock" was erroneously included in the Low GPC + High Rime condition. However, this item met criteria for the High GPC + High Rime condition, and was subsequently added to that condition. Thus, this resulted in 9 items in the Low GPC + High Rime condition, and 11 items in the High GPC + High Rime condition.

Irrelevant characteristics of the conflicting written word labels were controlled for across conditions in several ways: Normative word frequency for the conflicting labels was examined at first, second, and third grade levels to control for word frequency across conditions, using the Zeno, Ivens, Millard, and Duvvuri (1995) child language corpus. The word conditions did not differ in the normative frequency of the interfering labels (High GPC + High Rime: $M = 141$, $SD = 159$; High GPC + Low Rime: $M = 91$, $SD = 193$; and Low GPC + High Rime: $M = 166$, $SD = 171$; $F(2, 27) < 1$, $p = .653$). Further, the number of letters in the conflicting labels did not differ across conditions (High GPC + High Rime: $M = 3.8$, $SD = .4$; High GPC + Low Rime: $M = 4.0$, $SD = .8$; and Low GPC + High Rime: $M = 3.8$, $SD = .7$; $F(2, 27) < 1$, $p = .721$). To preclude the initial phoneme from affecting picture naming (Marmurek, Proctor, & Javor, 2006), none of the interfering labels began with the same initial phoneme as the to-be-named picture.

The characteristics of the picture stimuli were also controlled across conditions. The word frequency of the picture name did not differ across conditions for first through third grade according to the Zeno et al. word frequency norms (Control: $M = 204$, $SD = 254$; Letter String: $M = 113$, $SD = 149$; Low GPC + High Rime: $M = 228$, $SD = 435$; Nonword: $M = 67$, $SD = 59$; High GPC + High Rime: $M = 182$, $SD = 254$; and High GPC + Low Rime: $M = 158$, $SD = 190$; $F(5, 54) < 1$, $p = .723$). The initial phoneme in the picture names also did not differ significantly in the rank phoneme occurrence in spoken English according to the Berndt et al. norms (Control: $M = 13.4$, $SD = 6.7$; Letter String: $M = 13.6$, $SD = 7.1$; Low GPC + High Rime: $M = 20$, $SD = 9.1$; Nonword: $M = 17.4$, $SD = 10.1$; High GPC + High Rime: $M = 10$, $SD = 5.3$; and High GPC + Low Rime: $M = 14.4$, $SD = 6.6$; $F(5, 54) = 2.06$, $p = .084$). The number of syllables in the spoken picture name also did not differ across conditions (Control: $M = 1.4$, $SD = .5$; Letter String: $M = 1.4$, $SD = .5$; Low GPC + High Rime: $M = 1.4$, $SD = .5$; Nonword:

$M = 1.5$, $SD = .9$; High GPC + High Rime: $M = 1.5$, $SD = .7$; and High GPC + Low Rime: $M = 1.6$, $SD = .5$; $F(5, 54) < 1$, $p = .980$).

The predictability of the vowel pronunciation was deliberately varied across conditions according to the Berndt, Reggia, and Mitchum (1987) norms (High GPC + High Rime: $M = .60$, $SD = .30$; High GPC + Low Rime: $M = .51$, $SD = .16$; and Low GPC + High Rime: $M = .05$, $SD = .03$; $F(2, 27) = 19.52$, $p < .001$), but the predictability of the vowel GPC in the High GPC + High Rime and High GPC + Low Rime conditions that allowed us to measure the rime unit interference was controlled for, $F(1, 19) < 1$. Predictable and unpredictable rimes were taken directly from words used by Treiman Mullennix, Bijeljac-Babic, and Richmond-Welty (1995), who determined rime unit predictability by using a computerized version of the *Merriam-Webster Pocket Dictionary*, while controlling for consonant-vowel-consonant (CVC) pronunciation and the orthographic neighbors of each CVC word (Treiman et al., 1995). Examples of the stimuli can be found in Fig. 1.

Procedure—Children were tested in a quiet location in their school. Each participant was first seated in front of the computer screen, and asked to hold the voice-operated microphone in hand while the experimenter read a set of instructions aloud from the computer screen. The instructions stated “Welcome to the picture naming task. In this task, you will first see a little cross in the middle of the computer screen like this: (+). Then, you will see a picture with some writing over it. For instance, you might see a picture of a baby with ‘desk’ written over it. Tell us what the picture is and pretend the writing isn’t even there. Just name the picture. Can you tell me what you are supposed to do?” Each participant was then asked to restate the instructions and was shown how to hold the microphone properly so the computer could record the response. If the participant did not understand the instructions, the experimenter repeated the directions until the child understood that the purpose was to say the name of the picture and ignore the written word. Once the participant responded correctly, the experimenter then administered the practice stimuli. Before responding to the practice stimuli, a sentence stating, “Now we’re ready to do some practice. Get ready to name the picture,” appeared on the computer screen and was read aloud by the experimenter.

Six practice stimuli, one stimulus from each condition, were then presented. Following the presentation of the practice stimuli, another sentence appeared that stated; “Now we’re going to do the real thing. Just do your best and that’s good enough for me.” Each participant was then given 60 experimental trials. Each of the 60 experimental and six practice stimulus pictures was presented only once, and were presented in random order. Every participant received a different random order.

For each trial, a fixation point appeared when the experimenter pressed a button on the serial response box followed a second later by the picture stimulus. Participants then named each picture stimulus by speaking aloud into the microphone. The accuracy of the participant’s response was recorded by the experimenter. Each response was coded as correct if it was an appropriate name for the picture. The response was coded as incorrect if the verbal response was not an appropriate name. If the participant’s voice did not trigger the microphone, when a noise other than the participant’s response triggered the microphone (background noise, cough, sneeze etc.), or when the participant made a noise other than naming the picture (i.e. vocalizing a pre-articulation response such as “tsk,” or merely talking on an unrelated topic), the experimenter recorded a mechanical error.

Design—Picture naming time measured in milliseconds (henceforth, ms) and error rate (percent incorrect) served as dependent variables. A mixed 3×6 design was used.

There were three grade levels (first, second, and third) as between-subjects factors and six levels of Stroop (High GPC + High Rime, High GPC + Low Rime, Low GPC + High Rime, Nonwords, Letter strings, and Picture Controls) as a within-subjects factor.

Results

To identify picture naming times and error rates that qualified for further analysis, the following steps were conducted on the raw data. First, although pictures from Cycowicz et al. were chosen for their nameability, we wanted to eliminate picture stimuli that were unknown to a large proportion of children in our particular sample. Pictures with more than 20% error rates at any grade were considered too unnamable to be included in the analysis and were eliminated from the data set. This caused the elimination of the picture “suitcase”, with its superimposed word “shall.” Second, all reaction time trials associated with incorrect picture naming were eliminated from the data. Then, all mechanical errors were eliminated from the data. Mechanical errors accounted for 4.80% of the data. Next, mean picture naming times and standard deviations were calculated for every participant in the study. Then, reaction time trials greater than 2 standard deviations above each participant’s mean and less than 200 ms were removed from the data as outliers. Approximately 14.7% of the data was removed as outliers. Finally, for each participant, averages were calculated for picture naming times and error rates for each condition at each grade level. These averages formed the basic data for subsequent analyses. Picture naming time was viewed as the primary dependent measure because error rates approached floor once the “suitcase” picture was removed (Picture: 4.4, 1.6, 1.0; Letter string: 3.9, 2.9, 3.1; low GPC + high rime: 7.0, 5.0, 3.9; nonword: 5.3, 4.3, 3.1; high GPC + high rime: 5.8, 5.0, 5.0; and high GPC + low rime: 8.6, 6.6, 6.5; for first, second, and third grade, respectively). Analyses using error rates as a dependent variable are not discussed in detail except where they directly contradict those of naming times. Picture naming times can be seen in Fig. 2.

A repeated measures analysis of variance was conducted to evaluate the effect of grade and Stroop condition on reaction time (ms) and error (% incorrect). For analyses using participant mean reaction times as the dependent variable, the within-subjects factor was condition with six levels of Stroop (Control, Letter string, Nonword, Low GPC + High Rime Unit Predictability, High GPC + High Rime Unit Predictability, and High GPC + Low Rime Unit Predictability). Grade was a between-subjects factor with three levels (first, second, and third). With reaction time as the dependent variable, the main effect for Stroop condition was significant, $F(5, 1390) = 25.54, p < .001$. The main effect of grade was also significant, $F(2, 278) = 7.99, p < .001$, and a nonsignificant interaction between grade and condition, $F(10, 1390) = 1.62, p = .096$. Using error (% incorrect) as the dependent measure, there was a significant main effect for Stroop condition, $F(5, 1390) = 18.15, p < .001$, and grade, $F(2, 278) = 4.14, p < .017$, but a nonsignificant interaction between these factors, $F(10, 1390) < 1, p < .615$. These analyses appeared to indicate that picture naming was slowed in the two high GPC and nonword conditions only compared to the picture and letter string controls. Picture naming was generally slower for first and second grade children than third grade children.

We were interested in the interference caused by sublexical units rather than that caused by the mere presence of letters per se. Thus, we considered the letter string control to be a preferable baseline control since it contained letter information but no reading unit information. Therefore, a repeated-measures analysis of variance was conducted involving five levels of Stroop condition dropping the picture control condition, treating the letter string condition as the control. This analysis yielded a main effect of Stroop condition, $F(4, 1112) = 16.20, p < .001$ on picture naming time. The main effect for grade was also significant, $F(2, 278) = 6.63, p = .002$. Again, the interaction was nonsignificant, $F(8, 1112) = 1.10, p = .362$. The number of errors children made were similar across Stroop condition, $F(4, 44) = 1.08, p = .377$, and

there was no interaction between Grade and Stroop condition, $F < 1$. Younger children made marginally significantly more errors than older children, $F(2, 278) = 2.78, p = .064$. However, the main effect for Stroop condition on picture naming time seemed to indicate differences among conditions in terms of the degree to which the labels slowed picture naming, so this omnibus analysis was followed up by partial ANOVAs testing the direct predictions of the large-units-first and small-units-first hypotheses.

If the large units-first hypothesis is correct, then we should find interference in conditions containing conflicting highly consistent rime unit labels compared to letter controls, particularly for younger readers. The large-units-first hypothesis was tested by comparing picture naming in High GPC + High Consistency Rime Unit, Low GPC + High Consistency Rime Unit, and Nonword conditions to the letter controls. As predicted, pictures containing the High GPC + High Consistency Rime Unit words and nonwords were named more slowly than letter string controls ($F(1, 278) = 19.86$ and $29.18, p < .001$, respectively). However, particularly important for tests of this hypothesis, Low GPC + High Consistency Rime words did not cause interference in picture naming compared to letter string controls, $F(1, 278) = 2.69, p = .10$. These findings are contrary to the large-units-first hypothesis.

If the small units-first hypothesis is correct, then we should find interference in conditions containing conflicting high predictability GPC labels compared to letter controls, particularly for younger readers. Thus, if this hypothesis is correct, then we should find children to be slower to name pictures having High GPC + High Consistency Rime, High GPC + Low Consistency Rime, and Nonword labels compared to letter string controls. As noted above, both High GPC + High Consistency Rime and nonwords were slowed compared to letter strings (both $p < .001$). Moreover, pictures High GPC + Low Consistency Rime labels also took longer to name than letter string controls, $F(1, 278) = 43.13, p < .001$. Results of these partial ANOVAs support the small-units-first hypothesis in that all conditions bearing labels containing high predictability GPC units showed interference compared to controls. Further, none of the Stroop condition $\times$ Grade partial ANOVA interactions approached significance, suggesting that the influence of GPC units was similar at all grade levels.

Finally, we tested whether pronounceable nonwords interfered with picture naming similarly to words containing high predictability GPC units (i.e., the word conditions producing significant interference in the above analysis) by performing simple planned contrasts between the mean reaction times for nonwords and the high GPC word conditions. The mean RT for nonwords did not differ from either of these two word conditions (High GPC + High Consistency Rime, $F(1, 278) = 1.68, p > .10$; High GPC + Low Consistency Rime, $F(1, 278) = 2.51, p > .10$). Thus, it appears that high predictability GPC units are accessed automatically in the processing of nonwords.

The main findings of the study support the small-units-first hypothesis by indicating that labels following high predictability GPC rules consistently produced Stroop interference. These findings were true regardless of grade, showing the importance of GPC in beginning and more advanced reading (see Fig. 2). The fact that words with high consistency rime units interfered with picture naming only when they also possessed high predictability GPC units suggests rime units are not automatized obligatorily, and, therefore, are not favored in early reading. Further, the fact that words following inconsistent rime-units produced such interference is wholly opposite the relation proposed by the rime unit hypothesis.

Experiment 2

The finding that words with high consistency rimes by themselves produced no detectable Stroop interference in Experiment 1 at any point in early elementary school was rather unexpected. One possibility is that most of the children in that study were past the phase of

using word analogies in decoding. According to Goswami (1993, 1999), rime use represents an early phase of word decoding, and largely disappears by second grade. If this is correct, then studying individual differences in decoding in first grade may capture children for whom rimes are more available in word decoding. In this experiment, we examined the generality of the finding that young readers experience Stroop interference only from labels containing high predictability GPC units. If beginning readers are more likely to use rimes in decoding than children more experienced with reading, it may be that a study focusing on first grade will discern that words containing highly consistent rimes cause Stroop interference in young children just learning to read. In this experiment, we carried out an analysis of individual differences in early reading skill as a way of determining whether rime units processed automatically by children who are differently skilled in learning to read.

Method

Participants—Participants were 115 first grade children attending schools in metropolitan areas in Georgia, one parochial (29%) and the other public (71%). All were native English speakers. Children ranged from 6 years, 1 month to 7 years, 10 months ($M = 6$ years, 11 month, $SD = 5$ months). Of these 52% were male and 48% female, 17% European American, 71% African-American, 5% Asian-American, 5% Latino, and 2% Other.

Apparatus, stimuli, and procedure—The apparatus, stimuli, and procedure for the Stroop task were identical to that described in the previous experiment.

Assessments of reading ability—Children were assessed using both the Sight Word Efficiency (SWE) and Phonemic Decoding Efficiency (PDE; nonword pronunciation) subtests of the Test of Word Reading Efficiency (Torgeson, Wagner, & Rashotte, 1999). This standardized assessment asks children to read as many of word and nonwords in a list as quickly as they can in 45 s. Concurrent validity estimates reported in the TOWRE test manual have a median of .91 in Grades 1–3. Alternate form reliabilities have a median of .97 in Grades 1–3. They were also assessed using the Wide Range Achievement Test (WRAT; Wilkinson, 1993) reading subscale, an untimed measure of word reading ability. The administration manual of the WRAT reports a median reliability coefficient of .91 and correlations with other tests of reading achievement of between .69 and .87. These assessments were counterbalanced such that approximately half the children received the Stroop task first and the other half received the reading assessments first.

So that an analysis of early reading skill could be carried out, the standardized scores of the children on the three assessments (TOWRE-SWE, TOWRE-PDE, and WRAT) were, first, ranked ordered. Then, these ranks were totaled and children were ranked on the basis of this total. Finally, children were divided into three skill groups on the basis of this order. The mean standard scores on each instrument for the children falling into each group can be found in Table 1.

Design—Picture naming times measured in ms and error rate served as the dependent variables. A mixed 3×6 design was used with three reading skills levels (low, medium, high) as between-subjects factors and six levels of Stroop (High GPC + High Predictability Rime, High GPC + Low Predictability Rime, Low GPC + High Predictability Rime, Nonwords, Letter strings, and Picture Controls) as a within-subjects factor.

Results

To identify picture naming times and error rates that qualified for further analysis, the following steps were carried out on the data. First, as before, we wanted to eliminate picture stimuli that were unknown to >20% of children in this particular sample. This caused the pictures

“suitcase,” “arrow,” and “broom” to be eliminated. Further, three children had error rates in excess of 20%, suggesting either that they did not understand the task, or had some unforeseen difficulty in identifying the pictures. These children too were eliminated from the data (one child from each reading skill group). Second, all reaction time trials associated with incorrect picture naming were eliminated from the data. Then, all mechanical errors were eliminated from the data. Mechanical errors accounted for 4.40% of the data. Next, mean picture naming times and standard deviations were calculated for every participant in the study. Then, reaction time trials greater than 2 standard deviations above each participant’s mean and less than 200 ms were removed from the data as outliers. Approximately 10.5% of the data was removed as outliers. Finally, for each participant, averages were calculated for picture naming times and error rates for each condition. Reaction time was viewed as the primary dependent measure because error rates approached floor for most children. Overall means for picture naming time can be seen in Fig. 3.

A repeated measures analysis of variance was conducted to evaluate the effect of word elements responsible for interference and grade on picture naming time in ms and error rates. For these analyses, the within-subjects factor was Stroop condition with six levels of Stroop (Control, Letter String, Nonword, Low GPC + High Rime Unit Predictability, High GPC + High Rime Unit Predictability, and High GPC + Low Rime Unit Predictability) and reading skill was a between-subjects factor.

Analyses were conducted with both picture naming time and error rate as dependent measures. Picture naming was viewed as the primary dependent variable because error rate means (in percents) approached floor (Picture: 7.6, 2.0, and 2.7; Letter String: 6.0, 6.1, and 1.7; low GPC + High Rime: 10.0, 5.8, 6.5; Nonword: 7.2, 3.8, 2.7; high GPC + High Rime: 9.5, 5.3, 4.9; High GPC + Low Rime: 13.4, 6.5, 5.7; for low, medium, and high skill participants, respectively). Thus, only analyses of error rates that directly contradict those of picture naming times will be reported here. With picture naming time as the dependent variable, the main effect for Stroop condition was significant, $F(5, 560) = 22.93, p < .001$. The main effect of skill was also significant, $F(10, 560) = 4.39, p < .001$, but there was a significant interaction between skill and condition, $F(2, 112) = 4.63, p < .001$. The form of this interaction seemed to suggest that this significant skill difference among Stroop conditions was attributable mainly to a growth in interference from letter string labels with skill, suggesting that letters are still in the process of being automatized for low skill first grade readers. This was borne out by an analysis comparing the naming times of letter string controls to picture controls, Stroop condition $\times$ skill $F(2, 112) = 9.72, p < .001$. See Fig. 3.

Because we primarily were interested in determining whether young readers have automatized word units at a sublexical level beyond letters, a subsequent 3 skill $\times$ 5 Stroop condition interaction eliminating the picture controls from the analysis was carried out. The main effect of skill remained significant, $F(2, 112) = 5.43, p = .006$. There was also a main effect of Stroop condition, $F(4, 448) = 8.39, p < .001$. However, the skill $\times$ Stroop interaction was not significant, $F(8, 448) = 1.85, p = .066$. This main effect of Stroop effect appeared to indicate that picture naming was slowed in the two high GPC and nonword conditions compared to letter string controls.

Finally, we tested whether pronounceable nonwords interfered with picture naming in a manner similar words containing high predictability GPC units. Simple planned contrasts comparing the mean reaction times for nonwords and the high GPC word conditions showed that picture naming times for nonwords did not differ from either of these two word conditions (High GPC + High Consistency Rime, $F(1, 112) < 1, p > .10$; High GPC + Low Consistency Rime, $F(1, 112) = 1.47, p > .10$). Thus, even for first grade children, it appears that high predictability GPC units are accessed automatically in the processing of nonwords.

Analyses of error rates are only potentially meaningful to the extent that they outright contradict those of picture naming times (i.e., they are significant in the opposite direction than would be expected on the basis of reaction time analyses), and they do not approach floor. The main effect for Stroop condition was significant, $F(5, 560) = 5.72, p < .001$, as was the main effect for skill, $F(2, 112) = 11.23, p < .001$. However, the interaction was nonsignificant, $F(10, 560) = 1.13, p = .340$. Results were similar after excluding picture controls from the analysis. A 3 skill $\times$ Stroop condition ANOVA yielded a main effect of skill $F(2, 112) = 8.47, p < .001$, that indicated that low skill readers made more errors than high skill readers, and a nonsignificant interaction between the two factors, $F(8, 448) = 1.19, p = .302$. There was also a main effect of Stroop condition, $F(4, 448) = 5.02, p = .001$, but it seemed to be of a different form with regards to the Low GPC + High Rime condition than that noted for picture naming times (which will be discussed later).

If the large units-first hypothesis is correct, then we should find interference in conditions containing conflicting highly consistent rime unit labels compared to letter controls, particularly for the young readers used in the current study. As before, the large-units-first hypothesis was tested by comparing picture naming in High GPC + High Rime Unit, Low GPC + High Rime Unit, and Nonword conditions to the letter controls. As predicted, pictures containing the High GPC + High Rime Unit words and nonwords were named more slowly than letter string controls ($F(1, 112) = 6.86$ and $11.22, p < .001$, respectively). However, contrary to this hypothesis, Low GPC + High Rime words did not cause interference in picture naming times compared to letter string controls, $F < 1$, although they did produce a significant 2.7% more errors than controls, $F(1, 112) = 11.22, p < .001$. Unfortunately, this finding of interference in terms of errors but not picture naming times was compromised by the existence a speed accuracy tradeoff between these two variables in this condition (r between error rates and picture naming times across participants = $-.234, p = .012$). Moreover, low skilled first graders showed no greater tendency towards this Stroop interference from low GPC + High Consistency Rime words in terms of errors than did high skilled first graders, $F(2, 112) < 1$. Consequently, although this finding of larger error rates produced for picture naming in the low GPC + High Rime condition over pictures with letter string controls might constitute evidence for the large-units-first view, it was not irrefutable.

If the small units-first hypothesis is correct, then we should find interference in conditions containing conflicting high predictability GPC labels compared to letter controls, particularly in first graders. Thus, if this hypothesis is correct, then we should find first graders to be slower to name pictures in all conditions having labels with high predictability GPC units compared to letter string controls. As noted above, both High GPC + High Rime and nonwords were slowed compared to letter strings. Moreover, pictures with High GPC + Low Rime labels also took longer to name than letter string controls, $F(1, 112) = 22.77, p < .001$. Results of these partial ANOVAs support the small-units-first hypothesis in that all conditions bearing high predictability GPC units showed interference compared to controls. Further, none of the Stroop condition $\times$ skill partial ANOVA interactions approached significance, suggesting that the influence of GPC units was similar at all skill levels.

In sum, the results of this study confirm those of the first experiment in that stimuli containing conflicting labels with high predictability GPC units produced Stroop interference in picture naming beyond that caused by letters, regardless of early reading skill level. Despite the fact that they were still in the process of fully automatizing letters, even low skilled readers showed interference from GPC units beyond that found for letters, suggesting that these units are available to cause interference in picture naming from early in the development of reading.

Experiment 3

The previous two experiments have found minimal evidence for the role of rime unit consistency in Stroop interference. It may be that the automatic activation of high predictability GPC units found in the previous experiments is something more typical of less mature readers. One possibility is that, as children gain more experience with word reading, the units engaged become larger, as purported by the small-units-first hypothesis. Because the pronunciation of rime units may be more predictable than GPC units, at least according to some accounts, these units may become increasingly automatized such they become the operative unit in word reading. Alternatively, it may be that GPC units are the operative automatized unit in word reading at all ages. The current study was a replication of the previous experiments with adults. If rime units become increasingly automatized with age, then one might expect to find a difference in the pattern of Stroop interference in adults than was found in the previous studies such that conflicting labels containing highly consistent larger rime units now cause interference in picture naming.

A secondary purpose to the present experiment was to investigate the strategic use of small versus large units as a function of task. That is, some tasks such as nonword reading directly require the use of decoding strategies for their completion. As noted by Brown and Deavers (1999), it may be that people have use of both types of information, but opt to use rime based strategies in some cases. In this study, a second task was included to examine the task-dependent nature of the use of these units in skilled readers. The Brown and Deavers (1999) task of reading of ambiguous nonwords having both potential rime and GPC unit pronunciations was borrowed to examine the strategic use of smaller versus larger units. If the size of the reading units employed is task-dependent, then adults may show differential use of GPC or rime units in the Stroop task and the nonword reading task.

Method

Participants—Thirty-four adults from the undergraduate introductory psychology research pool at the University of Georgia participated in this study for course credit. Their ages ranged from 17 to 22 years ($M = 19.59$, $SD = 1.26$). There were 17 male subjects (50%) and 17 female subjects (50%). All subjects were native English speakers. The WRAT3 Reading test (Blue form) was administered to be certain that all participants were proficient readers. All participants read at or above a high school level (standard score 93 for ages 17–19 and 90 for ages 20–24) on the test to be included in the study. The participants all met this requirement and most exceeded it to score at a post-high school level (standard score $M = 112$, $SD = 6$).

Apparatus and stimulus materials for Stroop task—The apparatus and stimuli for the Stroop task were the same as used in Experiments 1 and 2. The participants were tested in quiet laboratory room at the University.

Stimuli and procedure for nonword naming task—For the nonword-naming task, a nonword list from Brown and Deavers (1999) was used to determine if they would use GPC or rime-units as a strategy to read unfamiliar words. The list was revised slightly to fit southeastern U.S. dialect pronunciations. The ambiguous nonwords were nonwords that could be pronounced differently depending on whether participants used a GPC rule strategy or a rime unit strategy. The consistent nonwords for which both the GPC and rime unit strategies would have produced the same pronunciation were used as fillers, as they provided no interesting contrast for pronunciations. These different types of nonwords were randomized to form one list from which participants were to read. The list was presented to the participants in 12 point black font on a white piece of 8 × 11 in. piece of paper attached to poster board for sturdiness.

Each participant was seated at a desk and the revised Brown and Deavers nonword list was presented to him or her. Participants were told, "Please read out loud from this list of nonwords. They're not real words; I just want you to tell me how they sound. You'll start at the top and go down the first column before starting on the next column. It is not a speed test, so take your time." The responses of the participants were recorded by the experimenter as using GPC, Rime, or "other" on a separate piece of paper. After the nonword list task was completed, participants completed the Stroop task.

Design—The Stroop task used a within-subjects design with six levels of the Stroop factor (High GPC + High Predictability Rime words, High GPC + Low Predictability Rime words, Low GPC + High Predictability Rime words, Nonwords, Letter Strings, and Picture Controls). Picture naming time measured in ms and error rate both served as dependent variables.

The Brown and Deavers nonword reading task used a within-subjects design having two levels of strategy (GPC, rime unit). The dependent variable was the percentage of ambiguous nonwords that were read either with GPC pronunciation or rime unit pronunciations.

Results

To determine the influence of unit size in automatized reading in comparison to strategic reading, it was necessary to analyze the two different tests in the study separately.

Nonword reading task—The pronunciations that the participants gave to the ambiguous nonwords were scored as GPC, Rime, or other. A GPC strategy response would pronounce the given vowel grapheme with the most common pronunciation (according to Venezky, 1970) with all consonants accurately pronounced. If the rime strategy is used, the pronunciation should rhyme with the analogous word for which its pronunciation is most consistent (e.g. pronouncing *dalk* to rhyme with *talk*) according to Brown and Deavers (1999). A response was scored as "other" if it conformed to neither of these pronunciation rules. For each participant, the proportion of rime pronunciations and GPC pronunciations were calculated.

All 34 adults without exception used rime units the majority of the time for ambiguous nonwords. GPC pronunciations were given for only a minority of ambiguous nonwords, while strategies that fell under "other" were used even less. The means and standard deviations for each condition can be found in Table 2. Thus, adults had a clear strategic bias towards rime pronunciation of ambiguous nonwords.

Stroop task—For the Stroop task, prior to analyzing picture naming times, it was necessary to eliminate problematic times. Incorrect picture naming times were eliminated from the data (.34% of the data). Mechanical errors and their corresponding picture naming times were also eliminated from the data (3.73% of the data). Next, outliers that were over 2 standard deviations above the mean for each participant were removed from the data (4.61% of the data). With these data points eliminated, mean picture naming times and error rates were calculated for each participant in each condition. Error rates were uniformly low (Control $M = 0$, $SD = 0$; Letter strings $M = 0$, $SD = 0$, Nonword $M = .01$, $SD = .029$; Low Predictability GPC with Highly Consistent Rime Units $M = .01$, $SD = .027$; High Predictability GPC with Highly Consistent Rime Units $M = 0$, $SD = 0$, High Predictability GPC with Inconsistent Rime Units $M = .01$, $SD = .024$) and at floor. Because of this, reaction time was considered the only dependent variable of interest. Figure 4 displays the mean picture naming times for each condition.

To determine the effect of word units on picture naming time (ms), a repeated measures analysis of variance was conducted comparing the time taken to name pictures in each of the stimulus condition's six levels (control, letter strings, nonword, high predictability GPC with highly consistent rime units, high predictability GPC with inconsistent rime units, low predictability

GPC with highly consistent rime units) with each other. The main effect for the Stroop condition was significant, $F(5, 165) = 14.12, p < .01$. Figure 4 suggests that there was more interference related to the two high GPC and nonword conditions than with the control, letter string, and low GPC condition.

One goal of the experiment was to compare the influence of GPC versus rime unit in automatic reading with adults. If large units (rime units) are operative for adults in automatic reading, then only conditions using high predictability rime units (Low GPC + High Rime, High GPC + High Rime) should show interference in picture naming over the letter string control alone. If GPC or small units were most influential in adult automatic reading, then the only conditions using high predictability GPC units (High GPC + Low Rime, High GPC + High Rime, Nonwords) should show interference in picture naming over letter strings alone. Alternatively, it may be that both unit types have been automatized in adults. In that case, all word types should show interference. To test these hypotheses directly, partial ANOVAs comparing each condition against the letter string control were carried out. Comparisons were made against the letter string control rather than the picture only control to be able to attribute the interference to word units per se and not the mere presence of letters alone. As in Experiments 1 and 2, naming times were no longer for pictures containing Low GPC + High Rime words than for random letter strings, $F(1, 33) = 1.59, p = .217$. However, pictures containing nonwords, High GPC + High Rime labels, or High GPC + Low Rime labels took significantly longer to name than pictures with letter strings, $F(1, 33) = 32.09, 8.30, 46.92$, respectively, all $p < .01$. The results of these partial ANOVAs point to greater interference with stimuli composed of highly consistent GPC units. High GPC + High Rime and High GPC + Low rime words as well as nonwords produced significant interference compared to letter string controls.

As in the previous studies, we determined whether pronounceable nonwords interfered with picture naming similarly to words containing high predictability GPC units. Simple planned contrasts between the mean reaction times for nonwords and the high GPC word conditions indicated that picture interference from nonwords did not differ significantly from either of these two word conditions (High GPC + High Consistency Rime, $F(1, 33) = 3.83, p > .05$; High GPC + Low Consistency Rime, $F(1, 33) = 1.05, p > .10$). Thus, as in the previous studies, it appears that high predictability GPC units were accessed automatically in the processing of pronounceable nonwords.

The two tasks in this study were designed to determine if adults utilize GPC or rime units when reading automatically, without intention, and when reading strategically out loud. Our findings point to the use of different sublexical word units depending on whether strategic reading is carried out or automatic reading. It was found that, with proficient readers, there was more interference with words containing highly predictable GPC units in the Stroop task, suggesting the operation of small units in automatic reading. When these same proficient readers were given a nonword-reading task, pronunciations indicated a definite preference toward rime units in strategic reading.

Both of these results, however, do collectively support the idea that what size units are most influential is task dependent (Brown & Deavers, 1999). This stresses the important role of both small and large size units in successful reading.

General discussion

The purpose of these studies was to determine the automatic use of large or small sublexical units in the absence of strategies that encourage their use. To this end, we created a Stroop task in which to-be-named pictures were presented with different types of conflicting labels superimposed on them. If children are reading words automatically, then it would be anticipated

that the characteristics of those words might cause them to differentially interfere with picture naming. The logic of the studies was that, by providing the children with the primary task of picture naming, interference patterns caused by the superimposed words would indicate the types of sublexical units that become automatized in early reading. By examining the interference caused by words containing rime units with consistent pronunciations to those containing rime units with inconsistent pronunciations, we should be able to ascertain whether large units are automatized. By comparing the interference caused by words containing GPC units containing highly predictable pronunciations to those containing GPC units with less predictable pronunciations, we should be able to ascertain whether small units are automatized.

In the first experiment, first, second, and third grade readers showed similar interference patterns. Words containing high predictability GPC units were more likely to interfere with picture naming than words with less predictable GPC units. Similarly, in the second experiment, first graders showed the same basic pattern occurred and the effect of rime unit consistency was not evident. First, second, and third graders, as well as first grade readers with differing skill all failed to show lexical interference from words with high consistency rime units if these words did not also carry high predictable GPC units. Because of the similarity of findings for children from first through third grade, in a third experiment, we looked at the other end of the skill spectrum by examining this type of Stroop interference in skilled adult readers. The same interference patterns were displayed. Again, GPC predictability was the more important factor in causing Stroop interference. Moreover, this Stroop interference occurred for nonwords containing these units as well in all experiments and at all skill levels.

Because of the general absence of rime unit consistency effects in Experiments 1 and 2, we wanted to learn whether a distinction between automatic and strategic processing could be made. To make this distinction, we asked adults to carry out an additional nonword reading task. In nonword reading, some decisions as to their pronunciations must be made, which allows for the operation of decoding strategies. When adults were asked to read ambiguous nonwords that had two potential pronunciations (a rime and a GPC pronunciation), virtually every single adult showed a preference toward reading nonwords using a rime pronunciation. However, in the Stroop task, most of these same adults showed a GPC pattern of interference. Thus, it appears that to a great extent the use of rime units is strategically-driven.

The current studies have implications for theories regarding the units that first support the link from speech to orthography as children learn to read. According to the large-units first view, the earliest phonological unit around which children develop metacognitive awareness is the syllable. The foundational basis for the development of syllable knowledge comes about through the phonological awareness children develop prior to learning to reading (Bryant, Bradley, MacLean, & Crossland, 1989; Stahl & Murray, 1994). Phonological awareness develops in a large-to-small unit progression through exposure to songs and books with rhymes which then later develops to include phoneme-level information (Adams, Foorman, Lundberg, & Beeler, 1998; Treiman & Zukowski, 1991). Prior evidence for the large-units first view has come primarily from studies using clue word procedures or training studies. Clue word procedures examine the influence of phonological rime primes (e.g., pain-rain) compared to other types of primes that share beginning spelling sequences on children's decoding of upcoming words. In general, but not always (Bowey, Vaughan, & Hansen, 1998; Wood, 2002), children are better able to decode upcoming words that carry matching rimes than when they do not (Goswami, 1993, 1999). Training studies have been carried out that teach children to segment words explicitly into onsets and rimes and examining the ease with which these versus other combinations of words units are taught (Bruck & Treiman, 1992). In each of these cases, children were found to benefit from rime unit treatments in contrast to other unit types. The current experiments suggest that the importance of these rime unit approaches may have some utility for children learning to read as one of the strategies that might be taught, because

skilled adults appear to use them when given the option and impetus to enact rime units for decoding. However, it is also equally clear from our studies that rime units may not typically serve as obligatorily processed automatized units for most children learning to read or for adults reading automatically.

According to the *small-units first* view, the earliest phonological unit involved in early reading is the phoneme, particularly because children are introduced to print, either informally, by being exposed to letter names (which generally carry their letter sounds), or formally by learning letter-sound correspondences (Duncan, Seymour, & Hill, 1997; Ehri, 1995; Stuart & Coltheart, 1988). These then give way to larger units as children become more proficient at recognizing positional, morphemic, and other context-sensitive sublexical units that become inferred from the lexicon and consolidated as units with growing expertise (Ehri, 1997). Evidence for this position tends to emerge from studies that show an increase in rime readings of nonwords with increasing reading experience in young children (Duncan et al., 1997), as well as an increasing ability to select out common rime unit elements with development (Duncan et al., 1997; Seymour, Duncan, & Bolik, 1999). Our studies agree with the small-units first view to the extent that we found that words with predictable GPC units are more likely to interfere in picture naming in early readers. However, we saw no increasing tendency for rime units to gain importance in automatic word reading with age.

Our findings are most consistent with a *flexible unit size* model (Brown & Deavers, 1999). According to this view, both unit types are potentially available for use fairly early in the development of reading (Stuart, Masterson, Dixon, & Quinlan, 1999) as learners develop a sight vocabulary. Most modern models of word reading in adults suggest that the lexicon contains information that pertains to both rime and GPC units (e.g., Brown, 1998), some of which seems to be driven by the characteristics of English spelling-sound mappings (Goswami, Ziegler, Dalton, & Schneider, 2003). According to this model, words units are applied flexibly and differentially according to the demand characteristics of the task. Thus, different tasks may highlight different sublexical units.

To this end, we think that a distinction like the one made by Duncan, Seymour, and Hill (2000) and Gombert (1992) may be necessary to account for the automatic versus strategic processing differences we noted in Experiment 3. According to Duncan et al., we need to distinguish between implicit or *epiphonological* awareness (the unconscious organization of sublexical units that occurs as part of normal language development, which is applied automatically) and *metaphonological* awareness (conscious reorganization or strategy that develops largely through the demand characteristics of the task or through direct instruction). Evidence for this strategic, metaphonological knowledge has been shown in a number of studies. Brown and Deavers (1999) for instance, noted that adults shifted the pronunciation of ambiguous nonwords towards rime pronunciations when the list contained a high proportion of rime-based items on it. Similarly, Goswami, Ziegler, Dalton, and Schneider (2003) compared the reading of three kinds of nonword lists: small-unit nonwords (regular GPC pronunciations), large-unit nonwords (rime pronunciations), and a mixed list that combined both types of words. They found that English adults performed more accurately and quickly on small-unit lists and large-unit lists than on the mixed list, showing what the authors called a “switching cost”. (German adults did not show this same effect because of the nature of written German.) This “switching cost” is the price paid for going back and forth between small and large-unit processing. These results support that in order to read successfully, English readers must eventually possess small-unit and large-unit strategies.

The finding of the Stroop tasks of Experiments 1, 2, and 3, however, suggested that the major unconscious, or epiphonological organizational feature, may be the GPC unit rather than rime unit. In every case, there was clear interference from labels containing high predictable GPC

units, but not necessarily from highly consistent rime units. Thus, we conclude that, when reading is automatic, influence from GPC units is most evident, but, in skilled readers at least, when reading allows for the use of strategies rime units may come into play.

The current studies have some limitations. First, one surprise for us was the uniformity of findings across the developmental spectrum with regards to the form and size of the Stroop effects. We had expected a general increase in Stroop interference with the development of reading skill. The existing literature did suggest that Stroop interference would eventually reach asymptote, but there was disagreement at exactly what point in the development of reading this might take place. For example, Stanovich et al. (1981) had placed the asymptote as early as the middle of the first grade year. Schadler and Thissen (1981) had placed the asymptote as late as fourth grade. However, Golinkoff and Rosinski (1976) had found no differences in Stroop interference between third and fifth grade. In our own studies, we tested children toward the end of the school year, even Experiment 2, which focused on first grade children. If the Stanovich et al. estimate is closer to being correct, we may have already had missed the point at which the shift in Stroop effect took place. Thus, future investigations should examine the development of the Stroop effect earlier in the process of learning to read than we did.

Second, it is unclear whether rime interference might be shown for automatic word reading if the study emphasized the presence of rime units to the ultimate extreme. That is, would it be possible to increase the availability of rime units in the Stroop task by increasing the proportion of high rime unit—low predictability GPC words used as labels? In the current study, high consistency rime units were present in two thirds of the word labels used in the Stroop task (as were high probability GPC's). Our best guess that it would not matter because our studies used a plurality of high consistency rime units.

Third, we did not collect information regarding the instruction practices of the schools from which the children came. These days, some popular early reading/ spelling programs make great use of strategies that emphasize “word families” or common rime units (e.g., *Making Words*, Cunningham & Cunningham, 1992), whereas others emphasize small units (*Reading Mastery V*, Engelmann, 1995). Deavers, Solity, and Kerfoot (2000) have reported contrasting use of rime versus GPC units in nonword pronunciation in children who have been taught to decode using small unit strategies versus rime unit strategies. This suggests metaphono-logical awareness, at least, can be altered by the instruction that the children receive. It is unknown whether all the children in our studies received small unit instruction in learning to decode and that is why they showed the GPC interference that they did. Frankly, we are unsure that instruction practice is likely to be a key explanatory factor given the consistency of our findings across children and adults, and general skills levels. Further, the children in our studies attended five different schools in three different cities and had a variety of teachers, materials, and approaches. Still, the role of instruction in producing patterns of automatic Stroop interference is an interesting one that deserves future attention.

In sum, the current experiments provide consistent evidence that interference in automatic word processing comes about through the operation of highly predictable GPC units. This evidence, taken by itself, suggests that GPC instructional strategies may be beneficial in early reading instruction.

Acknowledgments

We thank Barbara Bradley, Carolyn Groff, Elisabeth Meisinger, Matthew Quirk, Franklin Turner, and Deborah Woo for their assistance in data collection. We thank Yael Cycowicz for providing us the digital versions of the pictures used in this experiment. This project was supported by the Interagency Education Research Initiative, a program of research managed jointly by the National Science Foundation, the Institute of Education Sciences in the U.S.

Department of Education, and the National Institute for Child Health and Human Development in the National Institutes of Health (NICHD NIH). Funding for this project was provided by NICHD NIH Grant 7 R01 HD040746-06.

References

- Adams, MJ. *Beginning to read: Thinking and learning about print*. Cambridge, MA: MIT Press; 1990.
- Adams, MJ.; Foorman, BR.; Lundberg, L.; Beeler, T. *Phonemic awareness in young children: A classroom curriculum*. Baltimore, MD: Brookes Publishing; 1998.
- Bakan P, Alperson B. Pronounceability, attentiveness, and interference in the color-word test. *American Journal of Psychology* 1967;80:416–420.
- Berent I, Marom M. Skeletal structure of printed words: Evidence from the Stroop task. *Journal of Experimental Psychology: Human Perception and Performance* 2005;31:328–338. [PubMed: 15826234]
- Berndt RS, Reggia JA, Mitchum CC. Empirically-derived probabilities for grapheme-to-phoneme correspondences in English. *Behavior Research Methods, Instruments, & Computers* 1987;19:1–9.
- Bowey JA, Underwood N. Further evidence that orthographic rime usage in nonword reading increases with word-level reading proficiency. *Journal of Experimental Child Psychology* 1996;63:526–562. [PubMed: 8954608]
- Bowey JA, Vaughan L, Hansen J. Beginning readers' use of orthographic analogies in word reading. *Journal of Experimental Child Psychology* 1998;68:108–133. [PubMed: 9503648]
- Brown, GDA. The endpoint of skilled word recognition: The ROAR model. In: Metsala, JL.; Ehri, LC., editors. *Word recognition in beginning literacy*. Mahwah, NJ: Erlbaum; 1998. p. 121–138.
- Brown GDA, Deavers RP. Units of analysis in nonword reading: Evidence from children and adults. *Journal of Experimental Child Psychology* 1999;73:208–242. [PubMed: 10357873]
- Bruck M, Treiman R. Learning to pronounce words: The limitations of analogies. *Reading Research Quarterly* 1992;27:374–388.
- Bryant PE, Bradley L, Maclean M, Crossland J. Nursery rhymes, phonological skills, and reading. *Journal of Child Language* 1989;16:407–428. [PubMed: 2760133]
- Chall, JS. *Stages of reading development*. Vol. 2. Fort Worth: Harcourt Brace College Publishers; 1996.
- Christensen CA, Bowey JA. The efficacy of orthographic rime, grapheme-phoneme correspondence, and implicit phonics approaches to teaching decoding skills. *Scientific Studies of Reading* 2005;9:327–349.
- Coltheart V, Leahy J. Children's and adults' reading of nonwords: Effects of regularity and consistency. *Journal of Experimental Psychology* 1992;18:718–729. [PubMed: 1385612]
- Comalli PE Jr, Wapner S, Werner H. Interference effects of Stroop color-word test in childhood, adulthood, and aging. *Journal of Genetic Psychology* 1962;100:47–53. [PubMed: 13880724]
- Cunningham PM, Cunningham JW. Making words. Enhancing the invented spelling-decoding connection. *The Reading Teacher* 1992;46:106–115.
- Cycowicz YM, Friedman D, Rothstein M, Snodgrass JG. Picture naming by young children: Norms for name agreement, familiarity, and visual complexity. *Journal of Experimental Child Psychology* 1997;65:171–237. [PubMed: 9169209]
- Deavers R, Solity J, Kerfoot S. The effect of instruction on early nonword reading strategies. *Journal of Research in Reading* 2000;23:267–286.
- Duncan LG, Seymour PHK, Hill S. A small-to-large progression in metaphonological awareness and reading? *The Quarterly Journal of Experimental Psychology A: Human Experimental Psychology* 2000;53:1081–1104.
- Ehri LC. Do words really interfere in naming pictures? *Child Development* 1976;47:502–505.
- Ehri LC. Phases of development in learning to read words by sight. *Journal of Research in Reading* 1995;18:116–125.
- Ehri, LC. Learning to read and learning to spell are one and the same, almost. In: Perfetti, CA.; Rieben, L.; Fayol, M., editors. *Learning to spell: Research, theory, and practice across languages*. Hillsdale, NJ: Lawrence Erlbaum Associates, Inc; 1997. p. 237–269.
- Ehri LC, Wilce LS. Does word training increase or decrease interference in a Stroop task? *Journal of Experimental Child Psychology* 1979;27:352–364.

- Engelmann, S. Reading Mastery V. Worthington, OH: SRA Macmillan/McGraw-Hill; 1995.
- Golinkoff RM, Rosinski RR. Decoding, semantic processing, and reading comprehension skill. *Child Development* 1976;47:252–258.
- Gombert, JE. Metalinguistic development. London: Harvester Wheatsheaf; 1992.
- Goswami U. Toward an interactive analogy model of reading development: Decoding vowel graphemes in beginning reading. *Journal of Experimental Child Psychology* 1993;56:443–475.
- Goswami U. Causal connections in beginning reading: The importance of rhyme. *Journal of Research in Reading* 1999;22:217–240.
- Goswami U, Ziegler JC, Dalton L, Schneider W. Nonword reading across orthographies: How flexible is the choice of reading units? *Applied Psycholinguistics* 2003;24:235–247.
- Guttentag RE, Haith MM. Automatic processing as a function of age and reading ability. *Child Development* 1978;49:707–716.
- LaBerge D, Samuels SJ. Toward a theory of automatic information processing in reading. *Cognitive Psychology* 1974;6:293–323.
- Logan GD. Automaticity and reading: Perspectives from the instance theory of automatization. *Reading and Writing Quarterly: Overcoming Learning Difficulties* 1997;13:123–46.
- Maclean M, Bryant P, Bradley L. Rhymes, nursery rhymes, and reading in early childhood. *Merrill-Palmer Quarterly* 1987;33:255–281.
- Manis FR, Szeszulski PA, Howell MJ, Horn CC. A comparison of analogy- and rule-based decoding strategies in normal and dyslexic children. *Journal of Reading Behavior* 1986;18:203–218.
- Marmurek HHC, Proctor C, Javor A. Stroop-like serial position effects in color naming of words and nonwords. *Experimental Psychology* 2006;53:105–110. [PubMed: 16909934]
- Monsell S, Taylor TJ, Murphy K. Naming the color of a word: Is it responses or task sets that compete? *Memory & Cognition* 2001;29:137–151.
- Perry C, Ziegler JC, Coltheart M. How predictable is spelling? Developing and testing metrics of phoneme-grapheme contingency. *The Quarterly Journal of Experimental Psychology* 2002;55A: 897–915. [PubMed: 12188519]
- Pritchard VE, Neuman E. Negative priming effects in children engaged in nonspatial tasks: Evidence of early development of an intact inhibitory mechanism. *Developmental Psychology* 2004;40:191–203. [PubMed: 14979760]
- Rayner K, Posnansky C. Stages of processing in word identification. *Journal of Experimental Psychology: General* 1978;107:64–80. [PubMed: 624917]
- Redding GM, Gerjets DA. Stroop effect: Interference and facilitation with verbal and manual responses. *Perceptual and Motor Skills* 1977;45:11–17. [PubMed: 905071]
- Rosinski RR. Picture-word interference is semantically based. *Child Development* 1977;48:643–647.
- Schadler M, Thissen DM. The development of automatic word recognition and reading skill. *Memory and Cognition* 1981;9:132–141.
- Seymour PHK, Duncan LG, Bolik FM. Rhymes and phonemes in the common unit task: Replications and implications for beginning reading. *Journal of Research in Reading* 1999;22:113–130.
- Stahl SA, Murray BA. Defining phonological awareness and its relationship to early reading. *Journal of Educational Psychology* 1994;86:221–234.
- Stanovich KE, Cunningham AE, West RF. A longitudinal study of the development of automatic recognition skills in first graders. *Journal of Reading Behavior* 1981;13:57–74.
- Stroop JR. Studies of interference in serial verbal reactions. *Journal of Experimental Psychology* 1935;18:643–662.
- Stuart M, Coltheart M. Does reading develop in a sequence of stages? *Cognition* 1988;30:139–181. [PubMed: 3215001]
- Stuart M, Masterson J, Dixon M, Quinlan P. Inferring sublexical correspondence from sight vocabulary: Evidence from 6- and 7-year-olds. *The Quarterly Journal of Experimental Psychology* 1999;52A: 353–366. [PubMed: 10371874]
- Tipper SP, Bourque TA, Anderson SH, Brehaut JC. Mechanisms of attention: A developmental study. *Journal of Experimental Child Psychology* 1989;48:353–378. [PubMed: 2584921]
- Torgeson, JK.; Wagner, RK.; Rashotte, CA. *Test of word reading efficiency*. Austin, TX: Pro-Ed; 1999.

- Treiman R, Goswami U, Bruck M. Not all nonwords are alike: Implications for reading development and theory. *Memory & Cognition* 1990;13:357–364.
- Treiman R, Mullennix J, Bijeljac-Babic R, Richmond-Welty ED. The Special role of rimes in the description, use, and acquisition of English orthography. *Journal of Experimental Psychology: General* 1995;124:107–136. [PubMed: 7782735]
- Treiman, R.; Zukowski, A. Levels of phonological awareness. In: Brady, SA.; Shakweiler, DP., editors. *Phonological processes in literacy: A tribute to Isabelle Y. Liberman*. Hillsdale, NJ: Lawrence Erlbaum Associates, Inc; 1991. p. 67-83.
- Venezky, RL. *The structure of English orthography*. The Hague, The Netherlands: Mouton; 1970.
- Walton PD, Walton LM. Beginning reading by teaching in rime analogy: Effects on phonological skills, letter-sound knowledge, working memory, and word-reading strategies. *Scientific Studies of Reading* 2002;6:79–115.
- Wilkinson, GS. *The wide range achievement test: Administration manual*. Wilmington, DE: Wide Range Inc; 1993.
- Wood C. Orthographic analogies and phonological priming effects. *Journal of Research in Reading* 2002;25:144–159.
- Zeno, SM.; Ivens, SH.; Millard, RT.; Duvvuri, R. *The educator's word frequency guide*. Brewster, NY: Touchstone Applied Science Associates; 1995.

Picture Control

Letter String Control.

Nonword

Low GPC+ High Rime

High GPC+High Rime

High GPC+Low Rime

Fig. 1.
Example of pictures used in the experiments

Fig. 2. Mean picture naming time as a function of Stroop condition and grade level for first, second, and third grade children in Experiment 1

Fig. 3.
Mean picture naming time as a function of Stroop condition and reading skill for first grade children in Experiment 2

Fig. 4.
Mean picture naming time as a function of Stroop interference condition for adults in Experiment 3

Table 1

Mean standard scores on reading assessments for first grade children at each skill level in Experiment 2

Skill		TOWRE Sight word	TOWRE Phonemic decoding	WRAT Reading
Low <i>N</i> = 38	<i>M</i>	93	89	97
	<i>SD</i>	9	8	10
	Range	75–111	79–111	75–118
Medium <i>N</i> = 40	<i>M</i>	109	103	111
	<i>SD</i>	8	7	12
	Range	93–126	91–118	90–136
High <i>N</i> = 37	<i>M</i>	120	117	122
	<i>SD</i>	8	9	11
	Range	107–146	106–137	99–151

Table 2

Number of items for which adults used rime or grapheme–phoneme correspondence rule strategy with the nonword reading task

Strategy	<i>M</i>	<i>SD</i>
Rime	7.41	.99
GPC	2.24	.96
Other	.38	.55

Maximum = 10