Orthographic Analogies and Early Reading:
Evidence From a Multiple Clue Word Paradigm

Robert S. Savage, Louise Deault, Julia Daki, and Julie Aouad
McGill University

Two experiments using a variation of the clue word analogy task (Goswami, 1986) explored whether children can make orthographic analogies when given multiple clue words, beyond the known effects of purely phonological activation. In Experiment 1, 42 children (mean age 6 years and 8 months) were first taught 3 “clue” words (e.g., fail, mail, jail) and then shown target words sharing orthographic and phonological rimes (e.g., hail), phonological rimes (e.g., veil), orthographic and phonological vowel digraphs (e.g., wait), phonological vowel digraphs (e.g., vein), or unrelated controls (e.g., barn). All word types were advantaged at posttest over unrelated controls. A small additional advantage for orthographic and phonological rimes over phonological rimes was evident in by-participant analysis. Finally, regression analysis showed a specific relationship between onset-rime phonological awareness and orthographic rime clue word task transfer. Experiment 2 replicated Experiment 1 with 30 children (M age = 7 years, 0 months) and added a distinct group of children taught multiple clue words sharing vowel digraphs (e.g., gait, main, maid). Results showed advantages for all words over unrelated controls and a small additional advantage for orthographic and phonological vowel digraphs over phonological vowel digraphs in the by-participant analysis. Overall, results suggest that some young children do have the ability to make orthographic analogies when given multiple exemplars but that most improvement in target word reading reflects purely phonological activation. Practical steps for identifying genuine analogy use in a subset of children are thus described.

Keywords: reading, analogy, children, phonology, orthography

For some time now, cognitive–developmental models of early reading have rested on the assumption that accurate word recognition is a crucial corequisite to skilled reading comprehension (see e.g., Ehri, 2005; Perfetti, 2003, 2007; Snowling & Hulme, 2005; for reviews). In turn, phonological awareness, the ability to reflect on or manipulate the sound elements of the language, appears to be necessary but not sufficient for the acquisition of word recognition skills in English (Byrne, 1998; Goswami & Bryant, 1990; Snowling, & Hulme, 2005) and many, if not all, other languages (Caravolas, Volin, & Hulme, 2005; McBride-Chang et al., 2005; Perfetti, 2007; Share, 2008; Ziegler & Goswami, 2005).

Beyond this general consensus, two sorts of phonology-based models of reading acquisition are discernible in the literature in English. Some models of word reading acquisition are based on small grapheme-to-phoneme correspondences (GPCs). These models emphasize analysis of all constituent phonemes of a syllable (e.g., back segmented into b–a–ck; Coltheart, Curtis, Atkins, & Haller, 1993). Some GPC-based theories emphasize that word reading acquisition is marked by qualitative changes in the representation of spelling knowledge of GPCs in words (Ehri, 1992, 2005; Savage & Stuart, 2006; Stuart & Coltheart, 1988).

Rime-based models of early reading acquisition (Goswami, 1993, 1999; Goswami & Bryant, 1990) emphasize analysis of a syllable into the initial consonant or consonants (sometimes referred to as “onsets”) and “rimes” (the sub-syllabic unit containing the vowel and terminal consonant[s] of a syllable). In the interactive analogy model first proposed by Goswami, (1993), onset-rime phonological awareness is held to drive early onset-rime-based orthographic analogy use in beginner readers. In analogy tasks, children are shown a word such as beak, are told its pronunciation, and then are asked to read words such peak or bean.1 On the basis

---

1 The term “analogy” usually refers to approaches to reasoning wherein the relationships between sets of concepts are shared but where the content of the concepts involved is not. It might thus be argued that the process of deriving a pronunciation of a word where both the relation and the content (e.g., spelling and pronunciation of eat in bead and peak) are shared across words should not be referred to as an analogy. The term “analogy” was first used to describe the speeded reading of words and pseudowords by literate adults (Glushko, 1979; Kay & Marcel, 1981). Soon after, studies indentifying children’s apparent sensitivity to orthographic rimes in reading-accuracy tasks were reported and early rime analogy by children was seen as the developmental precursor of adult rime analogy shown in priming effects in speeded word-reading tasks (e.g., Goswami, 1986). No fully articulated conceptual model bridging these adult and developmental usages of the term “analogy” has yet been advanced beyond the shared role of larger rime units. Despite these various conceptual issues surrounding the term, “analogy” is now widespread in all of these literatures and for this reason alone is retained here.
of this and other evidence (e.g., Zorzi, Houghton, & Butterworth, 1998a 1998b), Goswami (1999) argued that onset-rime phonological awareness comes to represent a distinct pathway to reading in English independent of spelling knowledge or use of GPC rules.

In more recent rime accounts, language-specific influences of orthographic rimes have been posited. Ziegler and Goswami (2005) reviewed evidence that rimes are more consistently represented than are GPCs in English. GPCs are, however, excellent guides to pronunciation in other “transparent” spelling systems, such as German or Italian. On this basis, they advanced a psycholinguistic grain size hypothesis, which argues, in essence, that small grain-sizes reflecting GPCs will be most useful for all readers in orthographies, such as German and Italian, containing many words that can be pronounced accurately using such GPCs. In English, which allows many exceptions to pronunciations derived from GPCs, other larger grains of correspondence between clusters of letters and phonological abilities at the level of onset-rime, alongside the use of analogies based on orthographic rimes to learn to read, may also be beneficial. Goswami, Ziegler, and Richardson (2005) make the strong claim that differences in spelling-to-sound consistency between English and other transparent orthographies result in the use of different orthographic grain sizes by beginning readers.

There is some evidence that children’s use of rime analogy is indeed limited in transparent orthographies such as Finnish or Dutch (Holopainen, Ahonen, & Lyttinen, 2002; van Daal, Reitsma, & van der Leij, 1994). Goswami et al. (2005) claimed that early sensitivity to orthographic rimes is also reflected in performance in orally administered phonological awareness tasks. In their study, English and German children in kindergarten (considered prereaders) and Grade 1 (considered readers) were asked to detect the odd word out in sequences such as bank, tank, pink, where the shared rime (-ank) is always represented consistently in spelling, compared with word sets such as good, hood, loud, where the rime (-ood) can be spelled in more than one way (e.g., good vs. food). English readers in Grade 1 were the only group to show advantages for orthographically consistent oddity words. This study was executed in a cross-sectional manner. With four distinct groups, matching on extraneous pupil variables is crucial. However, little matching on important pupil background was reported, and the German and English children also differed in age. The best test of the hypothesis would be with two groups of age-matched children followed longitudinally as they move from preliteracy to becoming readers. Furthermore, the two experiments in the study did not in fact find the crucial three-way Consistency (consistent vs. inconsistent oddity stimuli) × Group (prereader vs. reader) × Language (English vs. German) interaction to reach statistical significance. There was also no direct assessment of the word reading ability for the oddity task items themselves, so the assumption that orthographic differences underlie phonological task differences cannot be verified. Together the evidence from this study should probably be treated very cautiously until data from well-executed replications are available.

However, even this broad dichotomy between rime and phoneme-based models of word reading acquisition is an oversimplification for several reasons. First, neither the rime nor GPC-based models are presented as exclusive accounts of word reading acquisition. The two models differ in the developmental sequence of children’s mastery of the two types of linguistic units. Rime-based models admit an important role for GPC units in learning to read once children have been taught and learned them, and similar distinct contributions of phoneme-to-grapheme correspondence knowledge in spelling that come to influence reading (Goswami, 1999). Many phoneme-based models admit the converse for later rime use, albeit after reliable GPC decoding has first been established (Duncan, Seymour, & Hill, 1997; Ehri & Robbins, 1992; Seymour, Duncan, & Bolik, 1999). The rime-based models have been termed large-units-first models, and the GPC-based models have been termed small-units-first accounts (G. D. A. Brown & Deavers, 1999).

Most small-units-first accounts argue that children become aware of the consistencies between spellings and sound at the level of orthographic rimes only after significant amounts of exposure to the spelling system (Treiman, Mullennix, Bijeljac-Babic, & Richmond-Welt, 1995). A good deal of evidence is consistent with this hypothesis (e.g., Bowey & Hansen, 1994; Bowey & Underwood, 1996; G. D. A. Brown & Deavers, 1999; Stuart, Masterson, Dixon, & Quinlan, 1999).

These small-units-first models can be contrasted with early rime use models that have argued that children’s very early sensitivity to phonological rimes, and a purported causal link between early phonological rime awareness and word reading, drives the use of orthographic rime at the beginning of word reading acquisition (e.g., Goswami, 1986, 1993; Goswami & Bryant, 1990). As noted earlier, the psycholinguistic grain size hypothesis appears to assume that both early and late rime strategies contribute to the emergence of sensitivity to orthographic rime consistency in English readers (e.g., Goswami et al., 2005; Ziegler & Goswami, 2005). There is some evidence that explicit onset-rime awareness in children aged 4–5 years is associated with their later sensitivity to rime units, when aged 8–9 years (Seymour et al., 2000). However, early and late rime-use strategies clearly reflect distinct cognitive mechanisms. Early orthographic rime use is based on very young children’s phonological sensitivity to rimes in the absence of much knowledge of the characteristics of English spelling system. Later rime use is based on sensitivity to orthographic rime consistency after substantial exposure to the orthography. It is the claim that children have the ability to use orthographic rimes early that is investigated here.

**The Strategic Use of Rime in Early Reading**

Although most models of reading described above differ substantially, they are all developmental accounts in the sense that they relate use of different phonological grains at the onset-rime and phoneme level for reading to chronological age and reading experience. Some subsequent models of early reading have instead emphasized a strategic approach to reading tasks by young children that is more dependent on the exigencies of the task than previous models assumed. G. D. A. Brown and Deavers (1999) argued that children’s early reading is characterized by a flexible-unit-size strategy. They argue that children experience a mismatch between strong rime phonological skills in early word reading alongside emerging letter-level spelling knowledge. As a result, children seek to develop both levels of representation in the mental lexicon. Consequently, if the demands of a particular experimental task allow children to use their phonological onset-rime skills more readily, they will operate at this grain, but if a task allows
children to use their orthographic letter-sound skills more readily, they will instead operate at this grain. Their approach to experimental tasks is in this sense strategic in that children choose flexibly how best to approach a given reading task with the phonological and orthographic resources they have available at that time. Consistent with this view, Brown and Deavers showed that when children are given a pseudoword, such as *dalk*, in isolation, pronunciation is more likely to reflect use of GPCs, whereas when cued by a concurrent rime-based analogy word, such as *talk*, pseudoword pronunciations of *dalk* will often rime with the word *talk*.

Some studies of pseudoword processing reporting both rime consistency and word regularity effects in reading are also broadly consistent with these claims of parallel grains of representation (Stuart et al., 1999). Recent studies of children’s self-reported strategy use in analogy tasks and isolated word reading tasks (Farrington-Flint, Coyne, Stiller, & Heath, 2008; Farrington-Flint & Wood, 2007) and studies showing differences in approaches to decoding following distinct forms of classroom instruction (Deavers, Solity, & Kerfoot, 2000) also suggest flexibility in phonological rime or phoneme unit size use in reading tasks. On the other hand, much evidence exists that on more explicit phonological tasks, prereading children show advantages for phonemes over rime (e.g., Cassidy & Smith, 2004; Geudens, Sandra, & Martensen, 2005; Savage, Blair, & Rvachew, 2006), suggesting that explicit phoneme and letter knowledge skills may not be mismatched in prereaders. Furthermore, as G. D. A. Brown and Deavers (1999) acknowledged, in experimental tasks where the clue word (e.g., *talk*) is available while children are asked to pronounce a target (e.g., *dalk*), the rime-based pronunciations reported might entirely reflect activation of purely phonological rime information (e.g., Savage, 1997, 2001).

Evidence for the Early Use of Rime in Reading Acquisition

One of the strongest sources of evidence for the early use of rime-based orthographic analogy in early reading comes from Goswami’s clue word analogy task. Here, children are first shown a set of analogy target words, such as *beak* and *bean*, and control words, such as *lake*. In a subsequent posttest, children are shown the pronunciations of clue words, such as *peak*, and reread the target words. Sometimes particular advantages in reading rime-analogous target words have been found at posttest, possibly suggesting a privileged status for onset and rime-based analogies (e.g., Goswami, 1986, 1988). For example, in a study by Goswami (1993, Experiments 2 and 3), prereaders and beginning readers read words sharing the rime (e.g., *wink-pink*), the onset (*trim-trout*), the head (*trim-trip*), or a vowel digraph *meat-heap*). Children made more analogies to words sharing onsets and rimes than to words sharing other letter clusters. Such results support the interactive analogy model of early onset-rime use (Goswami, 1993), in which phonologically underpinned onset-rime representations emerge first in reading development. Subsequently, several studies have reported equal transfer for rime and nonrime units in young children (Bowey, Vaughn, & Hansen, 1998; Duncan et al., 1997; Savage, 2001; Savage & Stuart, 1998), which does not so readily support the interactive analogy model.

Phonological Priming Effects in Clue Word Analogy Studies

Other research has suggested that variation in the clue word task influences evidence of clue word facilitation (G. D. A. Brown & Deavers, 1999; Muter, Snowling, & Taylor, 1994; Savage, 1997). The availability of clue word information at posttest appears to be associated with facilitation at posttest. The possibility that clue word information may act in some way to activate or prime only phonological information and not spelling information among words has been the focus of particular attention in the literature. Phonological priming was first reported in the adult reading literature (e.g., Tanenhaus, Flanigan, & Seidenberg, 1980) and refers to the activation of related pronunciations following exposure to a word pronunciation. In the developmental literature on analogy, “phonological priming” has come to be used as a general term to describe any evidence of the activation of rime phonology in the absence of relevant orthographic information in rime tasks that purport to measure orthographically based reading processes. Rime phonological priming is thus clearly indicated in the clue word task when (a) children are provided with only phonological clue word information at posttest and yet still show improvement in pronouncing target words (e.g., Savage, 1997) or (b) when showing transfer to target words that share pronunciations but not spelling patterns (e.g., *beak-see*; e.g., Bowey, 1999).

Many studies to date have reported very strong phonological rime-priming effects in the clue word task (Bowey, 1999; Bowey et al., 1998; Nation, Allen, & Hulme, 2001; Roberts & McDougall, 2003; Savage, 1997; Savage & Stuart, 1998; Wood, 2002; Wood & Farrington-Flint, 2001). This suggests that maybe very little or even none of the rime analogy effect reflects any learning about spelling patterns in words (cf. Farrington-Flint & Wood, 2007). Although aspects of methodology and interpretation of some reported effects have been disputed (e.g., Goswami, 1999), the findings and procedures used have been replicated and generalized across variations in clue word paradigms (Bowey, 1999; Nation et al., 2001; Savage, 2001). Perhaps decisively, Roberts and McDougall (2003) used an analogy task derived closely from one of Goswami’s own studies (Goswami, 1990). Here children were invited to read words sharing orthographic rimes (e.g., *most-post*), phonological rimes (e.g., *most-toast*), or ambiguous words (e.g., *most-cost*). Significant advantages for words sharing orthographic rimes over the other two word types would suggest that the analogy effect does have an orthographic component. Roberts and McDougall found, however, that phonological rimes were equivalent to orthographic rimes at posttest. This suggests strongly that improvement in target word reading in the clue word analogy task reflects the activation of only phonological information.

Studies of Multiple Clue Words

Given that phonological priming effects may explain all or a very large part of the pattern of improvement in target word reading in traditional clue word tasks, Savage and Stuart (1998, Experiment 2) sought to explore the effects of teaching multiple clue words and then asking children to read target words in a separate session where clues were not present. This approach may be a good way to encourage the use of genuine orthographic analogy as evidence from studies of children’s use of analogy in
inductive reasoning tasks suggests that multiple examples may increase analogy use (A. L. Brown & Kane, 1988; Farrington-Flint, Wood, Canobi, & Faulkner, 2004; Goswami, 1993). The presence of multiple clue word exemplars is also likely to encourage orthographic rime use. The use of multiple clues also more closely approximates the approaches that educators might use in classroom reading interventions using rime family approaches (e.g., Christensen & Bowey, 2005; Comaskey, Savage, & Abrami, 2009; Johnston, 1999). In sum, there are several reasons to believe that experiments with multiple clue words will have an increased sensitivity for detecting the use of orthographic rimes over experiments with single clue words.

Savage and Stuart (1998, 2001) pretaught children families of three rining words such as *bark*, *mark*, and *park*, and then asked to read words such as *dark* presented in isolation a few minutes later. Children showed that they could use pretaught clue word information to read analogous target words but were no better at this for rime-analogous words than vowel-analogous words (e.g., children were pretaught *bark*, *mark*, *dark*, and then shown the untaught word *part*). Savage and Stuart (2006) explored the longitudinal importance of the analogy strategy in a sample of 6-year-olds. Results demonstrated that vowel-based, but not rime-based analogy at age 6 predicted reading skills at age 8. It is important to note, however, that no control was provided for the possible effects of phonological rime priming in any of these studies. It is arguable that controls for a phonological prime are required to establish whether orthographic learning has occurred, even in this multiple clue word design, as priming effects may affect results in any context where rhyming words are presented over relatively short periods of time (Farrington-Flint & Wood, 2007; Roberts & McDougall, 2003). Study 1 therefore investigates this important unresolved issue contrasting target words sharing both orthography and phonology (e.g., *seek-reek*) and those sharing only phonology (e.g., *seek-beak*).

The Contribution of Rhyme and Phoneme Awareness to Analogy Use and Reading

Goswami and Bryant’s (1990) original model of reading acquisition makes two claims: First, children’s ability to recognize rimes in words is causally linked to the development of reading skills; second, that rime-based analogy represents the mechanism by which the rime-reading association is operationalized. Both questions have been addressed by exploring individual differences in reading and phonological skills. Regarding the first assertion, several studies demonstrated that rime-detection ability is indeed a unique predictor of reading (e.g., Bryant, Maclean, Bradley, & Crossland, 1990). More recently, there has been lively debate about the appropriate methodology for establishing specific rime-reading associations (e.g., Bryant, 1998, 2002; Castles & Coltheart, 2004; Hulme et al., 2002; MacMillan, 2002; Savage, 2001). It has also been argued by some that rime and phoneme skills are part of a single phonological ability, rather than distinct abilities (Anthony & Lonigan, 2004; Papadopoulos, Spanoudis, & Kendeou, 2009), although other researchers have found different patterns (Savage & Rvachew, 2006).

One specific methodological issue that has emerged from this debate is that, to be comparable, measures of phoneme and rime awareness must control for extraneous task demands. That is, all aspects of the task apart from the size of the phonological unit examined should remain equal, (Bryant, 1998). Under such conditions, Hulme et al. (2002) and Savage and Carless (2005) have shown that measures of phonemic awareness were good longitudinal predictors of reading skills, whereas onset-rime awareness was a weaker or nonsignificant predictor. This specific methodological issue also applies to Goswami and Bryant’s (1990) second claim that onset-rime phonological awareness helps children to make predictions about consistency of letter strings. Although several studies support Goswami’s assertion that there is a specific link between individual differences in children’s rime sensitivity and their early analogy use (e.g., Goswami, 1990b; Goswami & Mead, 1992), several others studies have found that phoneme-level skills predict analogy use in the clue word task (e.g., Savage & Stuart, 2001; Walton, 1995; Wood, 2000). All such studies are, however, limited by the fact that they used phonological tasks that were not equivalent in terms of the extraneous task demands. Most often, rime oddity tasks in which children judge the odd one out in a spoken set, such as *pin, win, sit, fin*, and in which they are not required to articulate or manipulate the shared phonological unit have been contrasted with tasks requiring articulation or manipulation of a phonemic unit, for example, in blending or segmenting phonemes (e.g., segmenting *cat* to make *c-a-r-t*) or in deleting a phoneme from within a word to produce a word or pseudoword (*What is beak without the b?*). Task demands and phonological unit size are thus perfectly confounded, rendering all results hard to interpret. In one of the few studies to use equivalent onset-rime and phoneme tasks (Roberts & McDougall, 2003), phoneme but not rime awareness predicted improvement in target word reading at posttest in the clue word task. Roberts and McDougall (2003) used a version of Goswami’s original clue word task and showed that the effects there reflected phonological priming. Thus, despite much research, it is still not clear what the relationship is between well-matched phonological tasks assessing rime- and phoneme-grains of representation and the use of genuinely orthographic analogy by young children. The present study therefore seeks to explore this issue.

The Present Research

This article thus addressed three questions:

1. Is there evidence for spontaneous orthographic rime analogy use, controlling for phonological priming effects?

2. Is there an advantage at posttest for target words sharing orthographic rimes, controlling for effects of priming and compared to target words sharing vowels?

3. What is the association between phoneme awareness, onset-rime awareness and vowel and improvements in target word reading at posttest?

Experiment 1 addresses Questions 1 and 3 by exploring the effects of teaching families of words sharing rimes on spontaneous reading of words sharing orthographic or phonological rimes or vowel digraphs. The study also explores the specificity of the link between individual differences in improvement in reading of target words sharing rimes and rime phonological awareness. Experiment
2 explores Question 2 by investigating the effects of teaching families of words sharing either (a) rimes or (b) vowel digraphs on reading of words sharing orthographic or phonological rimes or vowel digraphs.

**Experiment 1**

**Method**

**Participants.** Approximately 80 consent request letters were sent out initially to request participation in the study. Forty-two first grade students (23 girls, 19 boys) from local suburban elementary schools took part in the study following parental informed consent. Their mean chronological age was 6 years, 8 months (range = 6 years, 1 month to 7 years, 4 months). All students whose parents granted consent were included in the study; thus, students were not preselected or screened prior to commencing the study. Students were assessed on their reading ability, vocabulary, and letter-sound knowledge to establish the typicality of the sample.

Reading ability was assessed using the Wide Range Achievement Test (WRAT-3; Wilkinson, 1993), which is a test of single word reading. The mean WRAT-3 standard score was 97.21 (SD = 11.04). The Letter Sound Knowledge task is an experim-ental task that requires students to identify the sound that is associated with each letter of the alphabet. Scoring is based on the total number of letter-sounds identified, for a maximum of 26 points. In the present sample, students identified a mean of 20.45 letter-sounds (SD = 5.05). In addition, students were assessed on their receptive vocabulary ability using the Peabody Picture Vocabulary Test (PPVT-III; Dunn, & Dunn, 1996). The mean PPVT standard score was 98.19 (SD = 10.63). These scores all confirm that a group of children had been sampled that was quite typical on these measures.

**Design and procedure.** The study used a fully within-subjects design, with all students participating in each condition of the experiment. Students participated in a total of five 20–30-min sessions as part of the experimental paradigm. In Session 1, students completed the three background ability measures described above. In Sessions 2–4, students engaged in the analogy tasks, with two different word sets presented per session. The three analogy task sessions were counterbalanced by order across students to vary the presentation of word sets. In Session 5, students completed two phonological awareness tests involving rime production and final consonant (coda) production. The rime task and the coda task were also counterbalanced with respect to order of presentation within sessions.

Administration of the analogy tasks followed a carefully standardized procedure, originally developed in a modified version of the clue word analogy task (Goswami, 1986), which was based on the presentation of a single and often concurrently available clue word. The modified task was based on one described by Savage and Stuart, (1998, 2001). This approach allowed children to learn multiple clue word exemplars that were not made available when target words were re-presented. In the pretest phase, the first word set was administered to students in random order. Students were asked to read the 16 target and filler words out loud one at a time. Students’ responses were recorded verbatim on the scoring sheet to analyze reading errors. In the teaching phase, students were asked to read the three clue words sharing rimes (e.g., fail, mail, jail) one at a time. If students were unable to read the word, then they were told the pronunciation of the word by the examiner, with no additional explanation or guidance, apart from general encourage-ment. Students read the clue words in random order until reaching a threshold of three correct pronunciations of each word. Examiners recorded the number of trials needed to learn the clue words to this criterion. After accomplishing three correct pronunciations of each clue word, the examiner then placed all three words together and read them out loud before putting them away. After the teaching phase, students took a quick break by engaging in an unrelated visual search task for 2–3 min to separate the two phases of the experiment. Following this, students were posttested on the 16 target and filler words one at a time, and their responses were recorded. Finally, students were asked to read the three clue words one more time to check whether they remembered them and could have used them to make inferences at posttest. This entire procedure was then repeated again with the second word set to complete the session.

Two phonological awareness tasks were also administered to students in Session 5 using the materials and following the procedure described in Savage et al. (2006). In both common unit tasks, students were given two words and asked to identify the sound that was the same in the two words. In the rime-production task, students were given words such as heart and cart and had to produce the rime unit, art, as the shared sound. In the coda-production task, students were given words such as dog and bag and had to produce the coda unit, g, as the shared sound. Both tasks were administered using pictures of each word to help students represent and remember the words. For each task, students were given three practice trials before administering the 10 test items. Feedback was given during the practice trials to ensure that students knew the type of response for which the examiner was looking; however, no feedback was given during the test phase. For each item, students were given multiple prompts (pointing to the picture, pausing, repeating the words) until they provided a response.

**Materials and stimuli.** As shown in Table 1, a set of six clue words and corresponding target words was developed. These words were carefully selected with the intention of choosing words that were comparable in terms of reading difficulty, word length and frequency in the English language. For each word set, there were 16 corresponding target words: two orthographic and phonological rime-primed words, two phonological rime-primed words, two orthographic and phonological rime-primed words, two phonological vowel-primed words, two control words, and six filler words. The two orthographic and phonological rime-primed target words shared the same rime sound and spelling patterns with the clue words (e.g., seek-reek), whereas the phonological rime-primed words shared only the rime sound but not the spelling (e.g., seek-beak). The orthographic and phonological vowel-primed words shared the same vowel sound and spelling with the clue words (e.g., seek-deem), whereas the phonological vowel-primed words shared only the same vowel sound with the clue words (e.g., seek-hear). Control words shared neither rime nor vowel sound patterns, nor grapheme-phoneme correspondence with the target words (e.g., seek-tart). They provided a control against general improvements in word reading ability following the teaching phase. The six filler words were also unrelated and
served only to make sure that there were an equal number of words sharing letters and sounds with clue words, as there were words not sharing these patterns.

The target word sets were initially matched across word type for overall orthographic structure. That is, wherever possible, words with consonant-vowel, digraph-consonant (CVC) structures (e.g., *beak*) were used in all conditions of the experiment. Sometimes it was not possible to find CVCs to fit all experimental categories of stimuli. Where the construction of sets required more complex structures be admitted (e.g., *whirl* or *brood*), we sought as far as possible to use equal numbers of items with such structures across word types. The target word types were initially matched using the Children’s Printed Word Database (Masterson, Stuart, Dixon, Lovejoy, & Lovejoy, 2003; Stuart, Dixon, Masterson, & Gray, 2003). This measure was based on all words in 685 books found in the reading records of a sample of children aged 5–7 years in schools in the United Kingdom. The written word measure provides a sensitive measure of children’s orthographic knowledge that is closely associated with performance on word reading tasks in a range of experimental and intervention studies (see e.g., Savage, Carless, & Stuart, 2003; Stuart et al., 1999). The measure was preferred for this aspect of its validity, alongside its relative recency compared with other more dated but still frequently used measures (e.g., Carroll, Davies, & Richman, 1971). A recent enlarged and updated database based on \( N = 1,011 \) books has been created (Masterson, Stuart, Dixon, & Lovejoy, 2010). Stimulus items were also matched using this most recent database. The mean frequency of target words per million in the 2010 word set was 12.17 (\( SD = 26.40 \)). Preliminary analyses of variance (ANOVAs) also confirmed that there were no significant differences between the five experimental word types on this measure of frequency, \( F(4, 44) = 1.09, MSE = 711.64, p = .38, n_s, \eta^2 = .09 \).

### Results

Following standard practices, by-item and by-participant databases were first constructed for pre- and posttest target word reading scores,\(^2\) clue word reading scores, and phonological awareness measures. As in previous studies (e.g., Goswami, 1988; Farrington-Flint & Wood, 2007; Savage & Stuart, 1998, 2001), the measure of target word reading at pretest and posttest was the correct articulation of the appropriate unit in target words (e.g., the number of rimes correctly articulated in rime analogous words and the number of vowels correctly articulated in vowel analogous words). This provides a more sensitive measure of improvement that allows units of different sizes, such as rimes and vowels, to be more fairly compared. Preliminary analyses using standard procedures (Tabachnick & Fidell, 2007), confirmed that neither data set showed any marked deviations from normality or evidence of outliers, and they were suitable for general linear modeling of repeated measures with ANOVA. The means and standard deviations for the target word reading for the by-participants data set are presented in Table 2. Inspection of these means suggested an improvement in reading of all four analogous target words at posttest, with little evidence of improvement for control words. There was some evidence of differential improvement within this set of words, with the largest improvement evident for orthographic and phonological rime-primed words.

Preliminary analyses indicated that there was no interaction between target word reading and the fully counterbalanced order of target word teaching sessions, confirming that there were no

---

\( ^2 \) By-item and by-participant analyses have complementary strengths: By-item analyses add confidence that overall results do not reflect the disproportionate impacts of a subset of participants, and by-participant analyses add confidence that overall results do not reflect the disproportionate impacts of a subset of items.

---

### Table 1

<table>
<thead>
<tr>
<th>Word set</th>
<th>Clue words</th>
<th>O + P Vowel-primed</th>
<th>P Vowel-primed</th>
<th>O + P Rime-primed</th>
<th>P Rime-primed</th>
<th>Controls</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>seek</td>
<td>deem 0</td>
<td>heat 51</td>
<td>reek 0</td>
<td>beak 14</td>
<td>tart 11</td>
</tr>
<tr>
<td></td>
<td>peek</td>
<td>heel 0</td>
<td>lean 5</td>
<td>meek 5</td>
<td>teak 0</td>
<td>fowl 3</td>
</tr>
<tr>
<td></td>
<td>geek</td>
<td>root 11</td>
<td>rude 16</td>
<td>tool 5</td>
<td>rule 24</td>
<td>yarn 0</td>
</tr>
<tr>
<td></td>
<td>cool</td>
<td>boot 38</td>
<td>nude 0</td>
<td>fool 19</td>
<td>yule 0</td>
<td>germ 0</td>
</tr>
<tr>
<td></td>
<td>pool</td>
<td>firm 0</td>
<td>curd 3</td>
<td>dirt 76</td>
<td>curt 0</td>
<td>torn 27</td>
</tr>
<tr>
<td></td>
<td>spool</td>
<td>whirl 16</td>
<td>churn 0</td>
<td>flirt 0</td>
<td>blurt 3</td>
<td>fleet 0</td>
</tr>
<tr>
<td>2</td>
<td>squirt</td>
<td>brood 0</td>
<td>crude 0</td>
<td>troop 19</td>
<td>group 59</td>
<td>sport 30</td>
</tr>
<tr>
<td></td>
<td>scoop</td>
<td>proof 14</td>
<td>prude 0</td>
<td>droop 5</td>
<td>croup 0</td>
<td>cloat 0</td>
</tr>
<tr>
<td></td>
<td>whoop</td>
<td>waif 0</td>
<td>vein 3</td>
<td>hail 0</td>
<td>veil 0</td>
<td>dork 0</td>
</tr>
<tr>
<td></td>
<td>snoop</td>
<td>bait 5</td>
<td>cane 5</td>
<td>pail 5</td>
<td>bile 0</td>
<td>faun 0</td>
</tr>
<tr>
<td></td>
<td>fail</td>
<td>soak 14</td>
<td>tone 0</td>
<td>moat 8</td>
<td>dote 0</td>
<td>worn 30</td>
</tr>
<tr>
<td></td>
<td>mail</td>
<td>moan 3</td>
<td>cone 14</td>
<td>coat 176</td>
<td>rote 0</td>
<td>void 0</td>
</tr>
<tr>
<td></td>
<td>jail</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

---

Note. \( \text{O + P} = \text{orthographic and phonological}; \text{P} = \text{phonological}. \)

---

195ANALOGIES FROM MULTIPLE CLUE WORDS
order effects. Data were therefore analyzed using a 2 (Test: pretest, posttest) × 5 (Word Type: orthographic and phonological rime primed, phonological rime primed, orthographic and phonological vowel primed, phonological vowel primed, control) ANOVA with repeated measures on test and word type. There was evidence of a significant chi-squared for the Mauchly Sphericity Test in the word type and Test × Wordtype interaction in the by-participants analysis only, so the corresponding Greenhouse-Geisser statistics are reported for these analyses. In all other cases, the sphericity-assumed statistics of analyses are reported. A preliminary analysis confirmed that there was a significant Test × Wordtype interaction by-participants, $F(3, 138) = 12.74, MSE = 1.47, p = .001, \eta^2 = .24$; by items, $F(4, 44) = 12.94, MSE = 4.26, p = .001, \eta^2 = .54$.

Post hoc analyses using Tukey’s honestly significant difference (HSD) test revealed that there was a significant difference between means at pretest as well as at posttest. In both the by-participant and by-items analyses, the pretest differences were due to the words in the orthographic and phonological rime-primed condition being read significantly more often than all or most other word types at pretest.

To better explore the improvement in reading by word type at posttest, analyses of covariance (ANCOVAs) were run on the posttest scores with repeated measures on word type and with the five corresponding pretest word type measures as covariates for both by-participant and by-item analyses. There was again evidence of a significant chi-squared for the Mauchly Sphericity Test for word type in the by-participants analysis only, so the corresponding Greenhouse-Geisser statistics are reported for these particular results. The main analysis revealed a main effect of Word Type × Participants, $F(3, 115) = 9.90, MSE = 2.71, p = .001, \eta^2 = .22$; by items, $F(4, 24) = 3.30, MSE = 9.84, p = .03, \eta^2 = .36$. There were no other significant effects. The adjusted means are shown in Figure 1. Post hoc contrasts of the adjusted posttest means by word type using Tukey’s HSD test revealed that there were significant advantages for all target words over the control words. There was also a significant advantage for the orthographic and phonological rime-primed words over the phonological rime-primed and both types of vowel-primed words in the by-participants analysis. A very similar pattern was evident to that reported above in the by-items analysis with one important exception: The mean difference between the phonological rime-primed words and orthographic and phonological rime-primed words escaped conventional significance in this analysis. Together, these results show that there are significant effects of purely phonological rime-priming, even in the multiple clue word task used here. It is important, however, that there was a significant difference between the orthographic and phonological rime-primed words and both the phonological rime-primed words and both of the vowel-primed word types at posttest even after controls for pretests differences. This finding is consistent with the idea that children can make genuine orthographic rime analogies.

Effect sizes for the means in these analyses were also calculated using Cohen’s standard equation for mean differences (Cohen, 1988), based on the formula (posttest − pretest) / (pooled pretest standard deviation). These effect sizes are presented below the word type means in Table 2. Cohen (1988) has described effect sizes as “small” when $d = 0.2$, “medium” when $d = 0.5$, and “large,” when $d = 0.8$. Cohen (1988) and others have also cautioned against an overly rigid interpretation of these scores across different behavioral contexts; nevertheless, such analyses are widespread in the literacy-acquisition literature in establishing the practical importance of significant findings. Interpretation of the means presented in Table 2 using Cohen’s broad criteria show that the improvement for all target words at posttest, except the controls, was large in all cases. The value added difference in effect sizes between orthographic and phonological rime-primed and

![Figure 1. Adjusted posttest means in Experiment 1 (by participants).](image-url)

**Table 2**

*Means, Standard Deviations, and Cohen’s Effect Sizes for Words in Experiment 1*

<table>
<thead>
<tr>
<th>Target words</th>
<th>Pretest</th>
<th>Posttest</th>
<th>Cohen’s effect size</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$M$</td>
<td>$SD$</td>
<td>$M$</td>
</tr>
<tr>
<td>O + P Vowel-primed</td>
<td>1.98</td>
<td>2.36</td>
<td>4.02</td>
</tr>
<tr>
<td>P Vowel-primed</td>
<td>1.60</td>
<td>2.01</td>
<td>3.50</td>
</tr>
<tr>
<td>O + P Rime-primed</td>
<td>2.50</td>
<td>3.09</td>
<td>4.81</td>
</tr>
<tr>
<td>P Rime-primed</td>
<td>1.69</td>
<td>2.09</td>
<td>3.64</td>
</tr>
<tr>
<td>Controls</td>
<td>1.55</td>
<td>1.90</td>
<td>1.69</td>
</tr>
</tbody>
</table>

**By-items**

<table>
<thead>
<tr>
<th>Target words</th>
<th>Pretest</th>
<th>Posttest</th>
<th>Cohen’s effect size</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$M$</td>
<td>$SD$</td>
<td>$M$</td>
</tr>
<tr>
<td>O + P Vowel-primed</td>
<td>6.92</td>
<td>4.17</td>
<td>14.00</td>
</tr>
<tr>
<td>P Vowel-primed</td>
<td>5.58</td>
<td>4.42</td>
<td>12.25</td>
</tr>
<tr>
<td>O + P Rime-primed</td>
<td>8.75</td>
<td>3.91</td>
<td>16.75</td>
</tr>
<tr>
<td>P Rime-primed</td>
<td>5.83</td>
<td>2.73</td>
<td>12.83</td>
</tr>
<tr>
<td>Controls</td>
<td>5.42</td>
<td>4.03</td>
<td>5.92</td>
</tr>
</tbody>
</table>

*Note.* $O + P =$ orthographic and phonological; $P =$ phonological. Maximum $n = 12$ (by-participants analyses), maximum $n = 26$ (by-items analyses).

---

3 Sphericity tests assess the equality or symmetry of the variances between levels of factors in repeated measures designs in ANOVA. Results from ANOVAs violating this assumption of equality cannot be trusted, as they undermine the validity of the analysis. Violations of sphericity can be detected using Mauchly’s W test. Where the W test is significant, it is standard practice to make corrections for violations of sphericity using the Greenhouse-Geisser procedure. The Greenhouse-Geisser procedure alters the degrees of freedom, thereby altering the significance value of the $F$ ratio, and allows legitimate interpretation of main effects and interactions (e.g., Howell, 1997).
phono-logical rime-primed words and orthographic and phonological vowel-primed and orthographic and phonolog-
ic vowel-primed and phonological vowel-primed target words.

The final set of analyses sought to explore the relationship between individual differences in transfer in the current multiple clue word task and different levels of phonological awareness ability and clue word knowledge. Stepwise linear regression models were used for each of the clue words types except the control words that had shown no improvement in the analysis above. The number of target words articulated at posttest was the dependent variable in each case. Following the procedure in previous studies (Goswami & Mead, 1992; Roberts & McDougall, 2003; Savage & Stuart, 2001), the respective number of words read at the pretest for each target word dependent variable was entered at Step 1, followed by either the common unit rime or common unit coda tests at Step 2. Preliminary inspection of analyses showed uniformly high statistical tolerance scores (all tolerance scores in excess of .95), indicating that there was no evidence of collinearity in these analyses. The results of these analyses are depicted in Table 3. The results are very clear in showing that although pretest target word reading score was, as expected, a strong predictor of posttest target word reading, coda common unit test performance was never a significant predictor of improvement in target word reading between pre- and posttest when entered at Step 2. In contrast, as shown in Table 3C, rime common unit test performance was a significant and unique predictor of improvement in target word reading between pre- and posttest, specifically for the orthographic and phonological rime-primed target words. This pattern of significant association was not evident for the phonological rime-primed target words in Table 3D.

As a further check on the specificity of these associations, all regression analyses in Table 3 were rerun with WRAT word-reading entered at Step 2, prior to entering phonological awareness measures at Step 3. Although WRAT word reading was a significant predictor of growth in target word reading at posttest, the results for phonological awareness measures were essentially un-
changed by this additional analysis: The rime common unit test still explained a significant unique 4% in improvement in reading of orthographic and phonological rime-primed words. This pattern of significant and highly specific associations in Table 3 is thus consistent with the idea of a specific link between individual differences in rime phonological awareness and use of orthographic rime analogy task (Goswami & Mead, 1992). Rime common unit test performance was also uncorrelated with coda common unit task performance ($r = .19, p = .23, ns$), consistent with the idea that the two tasks tap into different forms of variance in phonological ability (Savage & Rvachew, 2006).

Last, inspection of the number of clue words read after the posttest was complete showed that children read a mean of 14.93 ($SD = 3.17$) clue words from a maximum of 18, showing that, for the most part, children had learned and had stored in their mental lexicons up to three relevant clue words for each analogous target word during the posttest. Hierarchical regression analyses were run to explore whether there was a significant relationship between pre- to posttest improvement in reading for each word type in the clue word task and the effectiveness of clue word learning. As with the analyses above, for the common unit tasks, posttest target word reading was the dependent variable, with respective pretest target word reading and then clue words known at posttest as the dependent variables at Steps 1 and 2, respectively. These analyses showed that clue word learning was associated with posttests improvements in target word reading in the rime-clued condition only, explaining a unique 4.5% of the variance in target word reading at Step 2.

**Discussion**

Experiment 1 sought to explore whether children can make genuine orthographic analogies following exposure to families of rime-analogous clue words, above and beyond the known effects of purely phonological priming. Results of both by-participant and by-item analyses provided clear evidence that children were able to make analogies to target words between pre- and posttest when using a multiple clue word paradigm in which the clue words were not present at posttest, replicating Savage and Stuart (1998, 2001). Beyond this, the results showed that children made inferences about pronunciations, even when the target words shared no or-

<table>
<thead>
<tr>
<th>Table 3</th>
<th>Regression Analyses Exploring Phonological Awareness as a Predictor of Clue Word Transfer</th>
</tr>
</thead>
<tbody>
<tr>
<td>Independent variable</td>
<td>Step</td>
</tr>
<tr>
<td>A. Orthographic and phonological vowel-primed target words (posttest)</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>2</td>
</tr>
<tr>
<td></td>
<td>2</td>
</tr>
<tr>
<td>B. Phonological vowel-primed target words (posttest)</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>2</td>
</tr>
<tr>
<td></td>
<td>2</td>
</tr>
<tr>
<td>C. Orthographic and phonological rime-primed target words (posttest)</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>2</td>
</tr>
<tr>
<td></td>
<td>2</td>
</tr>
<tr>
<td>D. Phonological rime-primed target words (posttest)</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>2</td>
</tr>
<tr>
<td></td>
<td>2</td>
</tr>
</tbody>
</table>

*p < .01. **p < .001.
thographic relationships. This effect is shown in the significant improvement in both the phonological vowel-primed and phonological rime-primed conditions in ANOVA, as well as in the robust effect sizes for improvement between pre- and posttest for both phonological vowel- and phonological rime-primed words. However, the finding that there was a modest but significant additional effect for orthographic and phonological rime-primed over phonological rime-primed words in the by-participants analyses is consistent with the idea that children also showed some significant capacity to make genuine orthographic inferences. In contextualizing this last result, analysis relevant to the real world importance of these findings showed that the Cohen’s effect sizes for orthographic and phonological rime-primed words were only marginally greater than that for phonological rime-primed target words, an additional value-added effect size that was small in size using Cohen’s (1988) standard criteria for measuring effect sizes.

Experiment 1 also sought to explore the relationship between individual differences in analogy use in the current multiple clue word task and different levels of phonological awareness ability, following the regression-based approaches reported in several existing studies (Goswami, 1990a; Goswami & Mead, 1992; Roberts & McDougall, 2003; Savage & Stuart, 2001). Results showed that the only significant association was between rime common unit task performance and improvement in reading of orthographic and phonological rime-primed words between pre- and posttest. This association is exactly the one first hypothesized by Goswami and colleagues (Goswami & Bryant, 1990; Goswami & Mead, 1992). This is, however, the first time that such a link has been demonstrated empirically using comparable rime- and phoneme-awareness tasks and in a clue word task using multiple clue words, controlling for the effects of phonological priming.

Experiment 1 also showed that children learned and had available a substantial amount of recently taught and relevant clue word knowledge. Results also showed that there was a significant association between individual differences in the extent of retention of clue word learning, as measured by the number of clue words read correctly after the posttest, and individual differences in improvements in reading in the orthographic and phonological rime-primed condition at posttest. This arguably provides further evidence that children used orthographic knowledge from taught clue words to read target words in this condition.

Overall, Experiment 1 has found evidence of rime priming and some limited but significant orthographic rime inference. One possible limitation of the experiment is that, throughout the study, children were only taught families of rimes, so this may have directed them toward rimes, possibly at the expense of other units such as shared vowel digraphs. Indeed, such a strategic approach is predicted from the flexible-unit-size hypothesis (G. D. A. Brown & Deavers, 1999; Farrington-Flint & Wood, 2007). Experiment 2 therefore sought to explore the patterns when children were exposed to families of clue words sharing vowel digraphs in comparison with exposure to families of clue words sharing rimes.

**Experiment 2**

**Method**

**Participants.** Around 90 consent request letters were initially sent out to request participation in the study. Thirty Grade 1 students participated in the second study (M age = 7 years, 0 months; range = 6 years, 5 months to 7 years, 7 months). Students completed the same background measures as in Experiment 1 to assess their reading, vocabulary, and letter-sound knowledge. Two children completed the pretesting but were unavailable for portions of the clue word testing sessions, with less than 30% of their data missing. Both of these children had typical reading and vocabulary scores. The missing clue word analogy data for these children was imputed using the SPSS missing value analysis. A preliminary analysis confirmed that data was missing completely at random (MCAR, Little’s MCAR), χ²(2) = 0.29, p = .86, ns, confirming the suitability of data for imputation. Data were then imputed using the expectation maximization (EM) procedure with five iterations. The mean reading level of the 30 participants who made up the final sample was 96.50 (SD = 14.58), as measured with the single word reading test from the WRAT-3 (Wilkinson, 1993). Students were also given the PPVT (Dunn & Dunn, 1996) and obtained a mean vocabulary standard score of 98.03 (SD = 12.56). In a test of letter-sound knowledge, students identified a mean of 21.40 letter sounds (SD = 3.24).

Independent of these preliminary assessments, students were randomly allocated within each classroom into one of two groups: a rime-taught group and a vowel-taught group. ANOVAs were then run to confirm that students’ reading, vocabulary, and letter-knowledge levels were the same across the two groups. A chi-squared analysis was run to confirm that there were equal proportions of boys and girls in both groups. The means for these groups and gender frequencies are depicted in Table 4. Independent t test analyses confirmed that there were no significant differences between rime-taught and vowel-taught groups on WRAT word reading, t(28) = 1.39, p = .18, ns; PPVT vocabulary, t(28) = 0.47, p = .64, ns; chronological age, t(28) = 1.16, p = .26, ns; or letter-sound knowledge, t(28) = 1.13, p = .27, ns. A chi-squared analysis of gender frequency confirmed that there were no differences in the gender balance of these groups, χ²(1) = .14, p = .71, ns.

**Design and procedure.** Unless otherwise stated, the method followed the same procedure as in Experiment 1. However, Experiment 2 had an important difference in that two randomized groups were created to teach students two different sets of clue words. In the rime-taught group, students were exposed to the

<table>
<thead>
<tr>
<th>Table 4</th>
<th>Means and Standard Deviations of Vowel-Taught and Rime-Taught Groups on Reading, Vocabulary, Letter-Sound Knowledge, and Gender</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td>Measure</td>
<td>M</td>
</tr>
<tr>
<td>WRAT-3 word-reading task</td>
<td>92.87</td>
</tr>
<tr>
<td>PPVT receptive vocabulary task</td>
<td>99.13</td>
</tr>
<tr>
<td>Letter sound knowledge task</td>
<td>20.73</td>
</tr>
<tr>
<td>Gender (boys)</td>
<td>9</td>
</tr>
</tbody>
</table>

Note. WRAT-3 = Wide Range Achievement Test (Wilkinson, 1993); PPVT = Peabody Picture Vocabulary Test (Dunn, & Dunn, 1996).
same material as in Experiment 1; that is, they learned clue words that shared the same rime unit. However, in the vowel-taught group, students were taught a new set of clue words that shared only the same vowel unit. These new words are shown in Table 5. Otherwise, for both groups, testing followed the same format as in Experiment 1. The presence of two groups each of 15 children that were taught different clue words precluded attempts to run regression analyses of individual differences in analogy use.

**Materials and stimuli.** The clue word sets used in Experiment 2 are depicted in Table 5. All stimuli for the rime-taught group remained the same; however, a new set of clue words was constructed for the vowel-taught group. These clue words had common vowel digraphs (e.g., **beef**, **sheep**, **weed**) but were not themselves rhyming words. Clue words for the vowel-taught group were carefully selected to be comparable to the words in the rime-taught condition with respect to word frequency and reading difficulty. As in Experiment 1, we again sought as far as possible to use equal numbers of items with CVVC structures across the both clue word sets. The mean frequencies of rime-taught and vowel-taught clue words per million in the present word set on the most recent version of the Children’s Printed Word Database (Masterson et al., 2010) were 31.06 (SD = 53.48) and 20.33 (SD = 38.88), respectively. There were no significant differences between these clue word groups on this measure of frequency, t(17) = .645, p = .53, ns.

**Results**

The preliminary data analyses were the same as in Experiment 1 and produced similar results, confirming the suitability of the data for mixed design ANOVA analyses. The means and standard deviations for the target word reading for the by-participants data set are presented in Table 6. Inspection of these means suggests an improvement in reading of all target word types at posttest, except the control words. The largest improvement evident was for the orthographic and phonological rime primed words.

### Table 5

**Table 5**

**Vowel-Based Clue Word Sets Used for the Vowel-Taught Group in Experiment 2**

<table>
<thead>
<tr>
<th>Word set</th>
<th>Clue words</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>beef, sheep, weed</td>
</tr>
<tr>
<td>2</td>
<td>zoom, spoon</td>
</tr>
<tr>
<td>3</td>
<td>hoop, chirp</td>
</tr>
<tr>
<td>4</td>
<td>mirth, smirk</td>
</tr>
<tr>
<td>5</td>
<td>gloom, spook</td>
</tr>
<tr>
<td>6</td>
<td>gait, main, maid</td>
</tr>
</tbody>
</table>

### Table 6

**Means, Standard Deviations, and Cohen’s Effect Sizes for Words in Experiment 2**

<table>
<thead>
<tr>
<th>Target words</th>
<th>Pretest</th>
<th>Posttest</th>
<th>Cohen’s effect size</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>M</td>
<td>SD</td>
<td>M</td>
</tr>
<tr>
<td>By-participants, rime-taught</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>O + P Vowel-primed</td>
<td>3.92</td>
<td>3.64</td>
<td>5.63</td>
</tr>
<tr>
<td>P Vowel-primed</td>
<td>2.43</td>
<td>2.72</td>
<td>4.64</td>
</tr>
<tr>
<td>O + P Rime-primed</td>
<td>3.47</td>
<td>3.14</td>
<td>6.19</td>
</tr>
<tr>
<td>P Rime-primed</td>
<td>2.59</td>
<td>2.45</td>
<td>5.03</td>
</tr>
<tr>
<td>Controls</td>
<td>3.34</td>
<td>3.22</td>
<td>3.26</td>
</tr>
<tr>
<td>By-items, rime-taught</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>O + P Vowel-primed</td>
<td>4.20</td>
<td>3.61</td>
<td>6.33</td>
</tr>
<tr>
<td>P Vowel-primed</td>
<td>3.47</td>
<td>3.66</td>
<td>5.07</td>
</tr>
<tr>
<td>O + P Rime-primed</td>
<td>4.87</td>
<td>3.98</td>
<td>6.60</td>
</tr>
<tr>
<td>P Rime-primed</td>
<td>4.13</td>
<td>3.58</td>
<td>5.73</td>
</tr>
<tr>
<td>Controls</td>
<td>3.73</td>
<td>3.62</td>
<td>3.60</td>
</tr>
<tr>
<td>By-items, vowel-taught</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>O + P Vowel-primed</td>
<td>4.00</td>
<td>2.34</td>
<td>5.67</td>
</tr>
<tr>
<td>P Vowel-primed</td>
<td>2.92</td>
<td>2.61</td>
<td>4.92</td>
</tr>
<tr>
<td>O + P Rime-primed</td>
<td>4.00</td>
<td>2.09</td>
<td>6.92</td>
</tr>
<tr>
<td>P Rime-primed</td>
<td>3.17</td>
<td>1.75</td>
<td>5.50</td>
</tr>
<tr>
<td>Controls</td>
<td>3.17</td>
<td>2.25</td>
<td>3.58</td>
</tr>
</tbody>
</table>

**Note.** O + P = orthographic and phonological; P = phonological. Maximum n = 12 (by-participants analyses), maximum n = 26 (by-items analyses).

Preliminary analyses confirmed that there was no interaction between target word reading and the order of target word teaching sessions, confirming that there were no order effects. Data were therefore initially analyzed using a 2 (Test: pretest, posttest) × 5 (Word Type: orthographic and phonological rime primed, phonological rime primed, orthographic and phonological vowel primed, phonological vowel primed, control) mixed design ANOVA with repeated measures on test and word type. In this study, there was evidence of a significant chi-squared for the Mauchly Sphericity Test in the word type and Test × Word Type interaction in the by-items analysis only, so the corresponding Greenhouse-Geisser statistics are reported for these analyses. In all other cases, the sphericity-assumed statistics of analyses are reported.

In the by-participants analysis only, this ANOVA analysis confirmed that although there was a significant Test × Word Type interaction, F(4, 112) = 9.40, MSE = 1.47, p = .001, η² = 25, post hoc analyses using Tukey’s honestly significant difference (HSD) test revealed there were also significant difference between means at pretest as well as at posttest. This time, the pretest differences were due to the words in the phonological vowel-primed and phonological rime-primed conditions being read significantly less often than the orthographic and phonological vowel-
primed word types at pretest. To better explore the improvement in reading by word type at posttest, ANCOVAs were run on the by-participants posttest scores with repeated measures on word type and with the five corresponding pretest word type measures as covariates. There was again evidence of a significant chi-squared for the Mauchly Sphericity Test in the word type and Test × Wordtype interaction in the by-participants analysis only, so the corresponding Greenhouse-Geisser statistics are reported for these analyses. The main analysis revealed a main effect of word type, \( F(3, 58) = 5.53, \text{MSE} = 4.14, p = .001, \eta^2 = .19 \). There were no other significant main effects. The adjusted means are depicted in Figure 2. Post hoc tests (Tukey’s HSD) showed that the main effect of word type was due to a significant difference between all target words and the control words and an advantage for the orthographic and phonological vowel-primed words over the phonological vowel-primed words.

Analyses of the by-items data using ANOVA revealed a main effect of test, \( F(1, 22) = 75.82, \text{MSE} = 2.52, p = .001, \eta^2 = .78 \), and a main effect of word type, \( F(4, 54) = 4.83, \text{MSE} = 12.58, p = .001, \eta^2 = .18 \). There was also a main effect of teaching group, \( F(1, 22) = 8.63, p = .008, \eta^2 = .28 \). In addition to these main effects, there was a significant Test × Word Type interaction, \( F(4, 51) = 7.75, \text{MSE} = 2.17, p = .001, \eta^2 = .26 \). There were no other significant interactions. Post hoc analyses using Tukey’s HSD test revealed that the control words again differed significantly from all other word type at posttest.

Effect sizes for the means in these analyses were again calculated. Interpreting results using Cohen’s (1988) criteria, the improvement for all target words at posttest except the controls was medium to large in all cases for the vowel- and rime-primed and clued words but were small for the control words. These are depicted in Table 6. The value added difference in effect sizes between orthographic and phonological rime-primed and phonological rime-primed words and orthographic and phonological vowel-primed and phonological vowel-primed target words derived from these mean scores was again small, \( d = -.13 \) and .08 versus .16 and .04 for orthographic and phonological vowel-primed and rime-primed words in the rime-taught versus the vowel-taught groups, respectively, in the by-participants analysis. For the same contrasts in the by-items analysis, the corresponding figures were -.14, and .24, versus .25 and .18.

Finally, to confirm that no confounds were introduced in the teaching of the vowel-taught and rime-taught words, independent sample \( t \) tests were run to confirm that the number of trials to learn words was the same. The means were 15.29 (\( SD = 11.28 \)) and 15.00 (\( SD = 9.78 \)), respectively, in the rime-taught and the vowel-taught conditions, with no differences evident between the number of trials to learn the word sets \( t(28) = .075, p = .94, \text{ns} \). In addition, the number of clue words recalled after the posttest across conditions was also analyzed. The means were 16.23 (\( SD = 1.09 \)) and 15.13 (\( SD = 2.48 \)), respectively, in the rime-taught and the vowel-taught conditions. Results again showed that there were no significant differences between the number of trials to learn the word sets, \( t(28) = 1.48, p = .15, \text{ns} \).

**Discussion**

Experiment 1 found evidence of phonological rime priming and some limited but significant evidence of orthographic rime inference. One possible limitation of the experiment is that throughout the study, children were only taught families of rimes, so this may have directed them toward rimes, possibly at the expense of other units, such as shared vowel digraphs. Experiment 2 therefore sought to explore the pattern of improvement in target word reading when children were exposed to families of clue words sharing vowel digraphs compared with exposure to families of clue words sharing rimes. Results of this study showed that there was again evidence of significant transfer following clue word exposure between pre- and posttest. Inspection of means showed that all words sharing spellings or pronunciations were advantaged at posttest over control words. There was no statistically significant advantage for the orthographic and phonological rime-primed words over phonological rime-primed words or for orthographic and phonological vowel-primed over phonological vowel-primed words in the by-items analysis. There was, however, some evidence of significant advantage for the orthographic and phonological vowel-clued words over the phonological vowel-primed words in the by-participants analysis only. There was no significant interaction whatsoever with teaching group (vowel-taught vs. rime-taught words), suggesting that although the shared units in multiple taught clue words do have a strong effect on improvements in target word reading at posttest, the presence of shared rimes versus shared vowels in taught words does not have a large influence on the relative levels of improvement witnessed at posttest to target words sharing vowels and/or rimes. Together, this suggests that the largest part of the pattern of improvement in target word reading might be best understood as reflecting phonological priming effects, rather than genuine orthographic inferences from taught words sharing vowels or rimes. Inspection of effect sizes confirmed this impression. Here, there were only very modest and occasional advantages for orthographic and phonological vowel-primed over phonological vowel-primed and orthographic and phonological rime-primed over phonological rime-primed words, and value added effect sizes were uniformly small. Beyond this, the modest advantage for orthographic and phonological vowel-primed over phonological vowel-primed words in some analyses shows that even young children can make orthographic vowel analogies (Savage & Stuart, 2006), and might suggest that children can be quite flexible in the orthographic unit used based on what clue word teaching has emphasized (G. D. A. Brown & Deavers, 1999).
General Discussion

This article addressed a series of research questions concerning the evidence for spontaneous orthographic rime analogy use. The first question concerned whether children in Grade 1 can show spontaneous use of orthographic analogy when taught clue words controlling for the known effects of phonological priming. This issue is important because although there is much evidence that children show improvement in target word reading at posttest in versions of the original clue word analogy task (Goswami, 1986, 1993), this may reflect phonological priming effects rather than true learning about orthographic rimes (e.g., Bowey et al., 1998; Roberts & McDougall, 2003; Savage, 1997; Savage & Stuart, 1998; Wood, 2002). The issue is also important because the influential psycholinguistic grain size hypothesis (Goswami et al., 2005; Ziegler & Goswami, 2005) suggests that early use of orthographic rime is one of the distinct characteristics of reading acquisition in English compared with German.

In the debate about priming effects in analogy tasks, however, there has been much less attention to tasks where children learn multiple clue words that are not present when target words are later presented. Among the few studies to explore this approach is the work of Savage and Stuart (1998, 2001, 2006). They were able to show that children do show signs of improvement in target word reading under such conditions but that patterns of improvement were equally strong for vowel-analogous as for rime-analogous target words. The present results replicated the pattern reported by Savage and Stuart (1998, 2001), as children again showed improvement in target word reading from pretaught clues not available at all to children in Grade 1, it is probably available to children to make rime analogies (Experiment 1), but where taught families of rime words, they extract orthographic rimes and use them to make rime analogies (Experiment 1), and where taught families of word sharing vowel digraphs, they can sometimes prefer to make vowel analogies (Experiment 2).

The current results instead suggest that phonological priming effects are pervasive in such experiments and make interpretation of learning in such tasks problematic. Such results might suggest to some that all versions of clue word learning tasks are thus flawed as measures of genuine orthographic inference use and that, as a consequence, researchers need to find new and more authentic ways to explore orthographic learning where priming artifacts are less relevant. If so, perhaps the most obvious way to do this is to direct more attention to training studies that directly compare rime- and phoneme-based teaching. There is, in fact, surprisingly little well-designed experimental research on this specific question (see e.g., Torgerson, Brooks, & Hall, 2006; Wyse & Goswami, 2008, for reviews; Comaskey et al., 2009; Savage, Abrami, Hipps, & Deault, 2009, for recent experimental intervention studies), and certainly more attention to intervention might well be warranted.

Although there is clear evidence of phonological priming in the present designs, a measured and balanced interpretation of the present results would also draw attention to the fact that there was some modest additional orthographic learning. These additional transfer effects were observed when target words shared both orthographic rimes and vowels in both Experiments 1 and 2 and sometimes reached significance in post hoc contrasts. The results of by-participant analyses suggest that where children are taught families of rime words, they extract orthographic rimes and use them to make rime analogies (Experiment 1), but where taught families of word sharing vowel digraphs, they can sometimes prefer to make vowel analogies (Experiment 2).

These results suggest that at least some children can make genuine orthographic analogies. A further case-by-case inspection of improvements made in Experiment 1 revealed that 16 of the 42 participants showed greater transfer for rime-clued than for rime-prompts. A similar pattern was evident in Experiment 2, where 13 of the 30 participants showed advantages for vowel-clued versus vowel-prompts. Withstanding issues surrounding measurement error, this analysis could be cautiously interpreted to indicate that if spontaneous orthographic analogy is available at all to children in Grade 1, it is probably available to only a subset of children.

In some respects, the finding of modest orthographic analogy effects in word reading outside of the original clue word task is consistent with other data. Children in Grade 1 report using an analogy strategy less than 5% of the time when reading isolated
words, whereas a phoneme-based decoding approach is four times more frequent in self-reports (Farrington-Flint et al., 2008). Stuart et al. (2003) and Masterson et al. (2010) analyzed rime frequencies in each of their corpora of all words in contemporary books for Grade 1 children. They both report that for the 300 most frequent CVC syllables in the database, there were 90 and 77 distinct rimes, respectively. Two thirds of these rimes occurred in only one word of the 300, suggesting that rime analogy is not likely to be a productive strategy for deciphering contemporary children’s texts.

Our final question examined the relationship between phonological awareness and analogy strategy use to explore what skills underpinned individual differences in analogy use. In addition to occasional main effects of orthographic analogy, Experiment 1 revealed that there was a significant and highly specific association between rime awareness, as measured using a common-unit task, and orthographic rime-based transfer. The existence of a strong link between these two skills has been long been suggested (e.g., Goswami, 1990b; Goswami & Bryant, 1990) and may be consistent with Goswami’s view that the use of rime analogy is the link between rime phonological skills and reading. The present study used matched explicit rime and phoneme (coda) awareness tasks for the first time in the multiple cue words task. It should be noted that explicit rime awareness shows a quite distinct developmental pattern compared with the implicit rime oddity tasks first used by Goswami and colleagues. For example, young children find common unit rime tasks harder than coda tasks (e.g., Duncan, Seymour, & Hill, 2000; Seymour et al., 1999), and prereaders generally cannot reliably complete rime common unit tasks at all, although they can show some modest facility with other units, such as codas (Savage et al., 2006). In contrast, there is much research showing prereaders; facility in rime oddity tasks.

Our interpretation of this specific orthographic rime strategy—rime phonological awareness link does reflect a genuine association between phonologically underpinned representations of well-specified orthographic rimes as Goswami (1993) first proposed, albeit underpinned by an explicit level of phonological awareness. However, the extent of orthographic analogy use in the present studies and the documented difficulty young children have with common unit rime tasks together suggests that at best, this yoked use of rime analogy and rime common units is a fledgling skill that is gradually emerging developmentally, rather than a cause of reading acquisition present in prereaders as Goswami and Bryant (1990) originally argued. As such, our results are consistent with the later use of rime units after significant reading experience proposed by several current models (e.g., Ehri, 2005; Perfetti, 2007; Savage & Stuart, 2006). Finally, the present study also found performance on the common unit rime and phoneme tasks was unrelated. This finding is consistent with the idea that the two tasks are measuring distinct phonological skills. Evidence of the specificity of the impacts of rime- and phoneme-based instruction on rime- and phoneme-awareness are also consistent with this general view (Comaskey et al., 2009).

Experiment 1 also revealed the existence of a link between the extent of clue word learning, as measured by the number of taught clue words correctly read after the posttest, and individual differences in transfer in the rime-clued condition at posttest. This also provides evidence that children used orthographic knowledge from taught clue words to read target words sharing rimes. This particular finding bears closer analysis, as it might suggest that the depth of learning of clue words is crucial to the use of early analogy. Ensuring depth of learning has been a problem for a long time in studies of early analogy use. Bruck and Treiman (1992) reported faster learning of rime pairs in their analogy study but poorer recall of these word pairs 1 day later. It may be, however, that with detailed knowledge of clue word orthography, even very young children might be able to make orthographic analogies. Bruck and Treiman highlighted the commonalities between word pairs with colored shared rimes in word pairs. It may be that even more detailed knowledge of words and perhaps their spellings, too, is needed before children can use clue words to map the pronunciations of sublexical units in related words they subsequently encounter (e.g., Ehri, 2005; Perfetti, 2007; Savage & Stuart, 2006).

Nevertheless, in a more general sense, there is other evidence that even young children start to learn the word-specific details of English orthography, Wright and Ehri (2007) showed recently that young children in kindergarten and Grade 1 learn words more effectively when they contain “legal” orthographic elements, such as doubled letters in coda, over “illegal” onset positions of words (e.g., JETT vs. RRUG), showing a nascent orthographic ability in this age group. Further work using more detailed and analytical clue word teaching may yet be repaid in evidence of early use of analogy.

Practically, how might educators identify children who do make genuine spontaneous analogies from those who do not? The present study suggests that the use of careful controls for phonological primes in the word sets used might be one way to do this effectively. Educators teaching children to read using rime families might routinely use foil items with shared rime pronunciation but distinct spelling (e.g., peek–meek with peek–beak) or other items with shared spelling and distinct pronunciations (e.g., read–bead and read–bread). The finding of specific associations between orthographic analogies and both explicit rime awareness and clue word knowledge after the posttest suggests other ways of accurately identifying children who make genuine orthographic analogies. A further case-by-case analysis showed that 86% of the children in Experiment 1 who scored zero on the common unit rime task showed no advantage in reading rime-clued compared with rime-primed words. Furthermore, of the subset of children who did show genuine orthographic analogy use in Experiment 1, all read at least 83% of the 18 taught clue words accurately after the posttest, whereas only a third of the children who showed no advantage for rime-clued over rime-primed words achieved this level of accuracy. Finally, the link between analogy use and accurate self-reports of analogy use reported by Farrington-Flint and Wood (2007) may further aid identification. From this view, children with good metalinguistic rime skills, who also retain accurate learning of clue words, and who can also articulate exactly how they read target words may be most likely to show genuine orthographic analogy use. Assessments of these three skills might be used together to identify children who can use orthographic analogy and to avoid misinterpreting phonological priming effects as orthographic analogy use in Grade 1 children.

**Limitations of the Present Study**

The present research has several important limitations. First, it is limited in scope to focus on the early use of rime- and vowel-based analogies in English and required the oral reading of isolated
monosyllables. The study is silent on other languages, or whether older readers can and do use rime or other forms of analogy. Similarly, findings reflect research on experimental studies of spontaneous analogy use and, as such, should be disconnected from studies explicitly teaching children to use rime analogies. Although the present results suggest that young children do not appear to be particularly adept at the spontaneous use of orthographic rime analogy, they may, however, benefit from being explicitly taught to make rime analogies, perhaps because of children’s phonological preferences for rimes in some tasks, or perhaps because rimes are consistent and may be useful patterns in the ultimate internalization of the English spelling system. Indeed, our own research (Di Stasio, Savage, & Abrami, 2010) provides some evidence of long-term advantage for rime-based early instruction, for a group of children from low-SES backgrounds with English as an additional language.

The present study did not feature analysis of the impact of rime-analogy teaching on spontaneous task performance, and there is some evidence that a heavy emphasis on rime- or phoneme-based approaches influences children’s decoding approaches (e.g., Deavers et al., 2000). Phonemic approaches did not feature heavily in our classrooms, where occasional reference to letters and sounds in a more holistic approach to early literacy was evident, rather than any systematic phonics teaching. Our results may well reflect these early educational experiences. Patterns of clue word task performance may however be different in different pedagogical contexts.

Implications of Findings and Future Research

The implications of the present findings for research and practice are several. First, for research, the results probably suggest that phonological priming effects are pervasive in clue word tasks. As a result, new research paradigms may be required that emphasize fuller and more explicit teaching of clue word spellings and rime-based decoding strategies, probably through well-designed intervention studies. The main implications for practice are cautionary in nature. Educational professionals should be wary of children’s apparent facility to generate pronunciations for target words on the basis of short-term learning of relevant clue words, as this approach may only reflect phonological priming. Educators can best verify whether children are using orthographic analogies appropriately by (a) contrasting pairs (e.g., peek-meek and peek-break) and (b) exploring children’s self-reported strategies (Farrington-Flint & Wood, 2007). These approaches to matching distinct student skills to learning strategies may maximize the chances of children using genuine spontaneous orthographic analogies in educational settings.

References


ANALOGIES FROM MULTIPLE CLUE WORDS


