Journal of Research in Childhood Education

Routledge
Taylor & Francis Group

ISSN: 0256-8543 (Print) 2150-2641 (Online) Journal homepage: https://www.tandfonline.com/loi/ujrc20

Sensitivity to the Regularity of Letter Patterns Within Print Among Preschoolers: Implications for Emerging Literacy

Quintino R. Mano & Heidi Kloos

To cite this article: Quintino R. Mano & Heidi Kloos (2018) Sensitivity to the Regularity of Letter Patterns Within Print Among Preschoolers: Implications for Emerging Literacy, Journal of Research in Childhood Education, 32:4, 379-391, DOI: 10.1080/02568543.2018.1497736

To link to this article: https://doi.org/10.1080/02568543.2018.1497736

Published online: 21 Aug 2018.

JOURNAL OF RESEARCH IN CHILDHOOD EDUCATION
2018, VOL. 32, NO. 4, 379–391
https://doi.org/10.1080/02568543.2018.1497736

Routledge
Taylor & Francis Group

Sensitivity to the Regularity of Letter Patterns Within Print Among Preschoolers: Implications for Emerging Literacy

Quintino R. Mano and Heidi Kloos

University of Cincinnati, Cincinnati, Ohio

ABSTRACT

When learning to read, the developing mind is likely to cluster letters into frequency-based chunks. In the current study, the authors investigated the extent to which such chunking takes place among preschoolers ($N = 54$) by examining the association between sensitivity to subword orthographic regularity and preschooler age. A version of the wordlikeness judgment paradigm was administered to assess sensitivity to the frequency of particular grain sizes (unigram, bigram, and trigram), regardless of their positional frequency within words. The task manipulation affected bigram performance but not unigram or trigram performance and was observed only among older preschoolers. Results suggest that preschoolers first become sensitive to bigram frequency. The special status of bigrams within subword frequency sensitivity, over and above unigrams and trigrams, has practical implication for reading instruction and remediation.

ARTICLE HISTORY
Received 16 May 2017
Accepted 16 March 2018

KEYWORDS
Literacy acquisition; orthography; preschool; statistical learning

The fact that hundreds of thousands of words are created from a limited number of letters is a fundamental feature of any alphabetic orthography. In English, for example, this fundamental feature produces a distributional phenomenon whereby some letter clusters (e.g., TH, ND, TIO, ATI, ATIO) within words occur more regularly than other letter clusters (e.g., UR, SI, ITI, AIN, ECTI). The number of letters forming a cluster can vary (e.g., bigram, trigram), with different grain sizes referred to as *n*grams. Some of these letter clusters may form words when presented alone (e.g., HE, IN, FOR, ATE) and can best be calculated as word frequency; however, because these letter clusters are chunks that can be recombined to form many words, they also may be thought of as belonging to the broader concept of subword regularity.

In recent years, with the development of freely available online databases, the frequency of *n*gram occurrence within the English language may be more easily calculated by researchers and educators. For example, the Children's Printed Word Database—an online resource—is a computerized database of words that provides a variety of orthographic and phonological properties of words that appear in books for children in the first two years of school. As another example, MCWord (Medler & Binder, 2005) is an online database that offers frequency-based calculations of *n*grams based on the CELEX database, which includes all the English wordforms from a COBUILD corpus of written and spoken text (i.e., approximately 16,600,000 written examples, and 1,300,000 spoken examples). MCWord (Medler & Binder, 2005) calculates statistics of *n*grams based on positionally constrained and positionally unconstrained frequencies. Other websites provide a rank order of subword structures (e.g., unigram, bigram, trigram), which are already calculated and made available for public use (Mano & Medler, 2017) . These online

CONTACT Quintino R. Mano quintino.mano@uc.edu Department of Psychology, University of Cincinnati, 47 Corry Boulevard, Edwards 1 Building, Suite 5130C, Cincinnati, OH 45219.

resources allow researchers and educators an opportunity to develop new protocols for reading instruction that emphasizes orthographic learning.

The distributional phenomenon of subword orthography places unique processing demands on the developing mind. These processing demands may stem from a natural learning propensity to detect patterns and statistical regularities within the environment, a propensity that is seen across the life span. For example, research shows that 8-month-old infants and adults are capable of detecting statistical properties of sounds presented within streams of linguistic and nonlinguistic stimuli (Saffran, Aslin, & Newport, 1996; Saffran, Johnson, Aslin, & Newport, 1999). Statistical learning of reoccurring visual patterns has also been observed among infants and adults (e.g., Kirkham, Slemmer, & Johnson, 2002; Turk-Browne, Jungé, & Scholl, 2005). Thus, the processing demand of subword distributions may be related specifically to statistical learning mechanisms. This is especially important for reading because statistical learning is linked to performance on standardized measures of word reading, observed in children and adults (Arciuli & Simpson, 2012).

The ability to extract statistical regularities within the environment reflects a powerful learning mechanism, which recent theoretical work suggests is equally applied to the textual environment (Mano, 2016). According to Mano (2016), the ability to extract regularities from the textual environment emerges from the dynamic interplay between incidental statistical learning and visual attentional control. The interplay between these processes, in theory, gives rise to sensitivity to the statistical regularities of subword orthography. It is conceivable that developing sensitivity to reoccurring *n*grams represents an important facet of typical reading development. Moreover, given that infants appear capable of detecting statistical regularities within auditory and visual modalities (Kirkham et al., 2002; Saffran et al., 1996), it is conceivable that developing sensitivity to regularities within the textual environment emerges from the earliest moments of print exposure. In fact, print exposure has recently been shown to contribute directly and indirectly to reading fluency (Mano & Guerin, 2018). We reasoned that the earliest moments of print exposure are those in which parents and/or guardians read to their children (e.g., bedtime reading) and during scheduled reading time in preschool. The overarching purpose of the present study was to examine the extent to which frequency-based chunking of *n*grams occurs among preschoolers.

To examine the earliest moments in which sensitivity to subword orthographic regularity may be observed, we investigated the association between sensitivity to subword orthographic regularity and preschooler age. We reasoned that older preschoolers will have been exposed to more print than younger preschoolers, and thus older preschoolers would be more likely to exhibit sensitivity to subword orthographic regularities compared to younger preschoolers. In the present study, an analysis of variance further examined the interaction between preschooler age and *n*gram regularity within print (discussed below).

*N*gram regularity within print

The distributional phenomenon of *n*grams within the textual environment is not a trivial factor of alphabetic orthographies. This is because accumulating evidence suggests that children develop sensitivity to the statistical regularity of letter patterns within print (Apel, Brimo, Wilson-Fowler, Vorstius, & Radach, 2013; Apel, Wolter, & Materson, 2006; Cassar & Treiman, 1997; Conrad, Harris, & Williams, 2013; Ise, Arnoldi, & Schulte-Körne, 2014; Kessler, 2009; O'Brien, 2014; Pollo, Kessler, & Treiman, 2009; Treiman, Kessler, Boland, Clocksin, & Chen, 2017; Wright & Ehri, 2007). For example, when shown the pseudohomophonic pair *hool* and *hewl*, school-age children are more likely to select *hool* as being more visually wordlike than *hewl* (Conrad et al., 2013). This is because *hool* comprises letter clusters that occur more frequently than the letter clusters that make up *hewl*. Sensitivity to the statistical regularities of letter patterns on similar wordlikeness judgment tasks has been observed as early as kindergarten (e.g., Apel et al., 2013). The development of sensitivity to the statistical regularities of letter clusters within orthography reinforces the notion that statistical learning mechanisms are applied to the textual environment (Arciuli & Simpson, 2012; Mano, 2016).

Evidence also indicates that the ability to detect reoccurring letter patterns within print explains unique and significant variance in multiple reading outcomes, including isolated word reading, rapid naming of letters, and reading fluency (Cassar & Treiman, 1997; Conrad et al., 2013; Ise et al., 2014; O'Brien, 2014; Pacton, Perruchet, Fayol, & Cleeremans, 2001; Stanovich & Siegel, 1994). These lines of research suggest that the ability to encode the distributional properties of subword orthography represents a unique skill within reading development. These lines of research also highlight the need to explore sensitivity to *n*gram frequency in early childhood, particularly in the preschool setting. Such explorations may show that sensitivity to *n*gram frequency is present from the earliest moments of literacy acquisition and may also yield important developmental insights.

Wordlikeness judgment: Assessing sensitivity to subword regularity

Perhaps the most common method for assessing sensitivity to subword regularity is the visual wordlikeness judgment paradigm (O'Brien, 2014; see for review, Mano, 2016). The typical format of this paradigm involves viewing a pair of pseudowords (e.g., *sloz* – *wosl*) that are manipulated in terms of the relative frequency of constituent letter chunks (the bigram *sl* is more common in the word-initial position than the bigram *wo*). The task is to select the item that looks most like a word. Note that the task is not to select the item that looks most like an existing word. Such a task would explicitly introduce an element of lexical referencing and thus expose the task to semantic confounds. Instead, the task is merely to make a judgment about which item within the pair looks like it could be a word. The basis on which these judgments are made is thought to be a result of implicit processing of wordlike features (e.g., *n*gram frequency, positional regularity of letters, conventional spelling patterns). Notably, the issue of whether these implicit processes are available for verbal report remains an important empirical question. To the best of our knowledge, no study to date has asked participants whether they based their wordlikeness judgments specifically on the statistical regularity of letter clusters or on some other property.

O'Brien (2014) showed elementary school children (Grades 1–3) *sloz* paired with *wosl* and instructed them to select the nonword they judged to be more wordlike. She found that children were more likely to select *sloz* than *wosl*, presumably because the constituent letter clusters within *sloz* occur more frequently in their spatial location than the constituent letter clusters within *wosl*. In other words, affirmative wordlikeness judgments in favor of *sloz* over *wosl* stems from the fact that the bigram SL is more likely to occur in the word-initial position than in the word-final position. Such sensitivity correlated significantly with performance on reading outcomes such as rapid naming. Moreover, those findings corroborate other studies showing associations between sensitivity to subword orthography and reading outcomes (e.g., Cassar & Treiman, 1997; Conrad et al., 2013; Pacton et al., 2001; Rothe, Schulte-Körne, & Ise, 2014; Stanovich & Siegel, 1994).

The wordlikeness judgment paradigm possesses versatility insofar as it can accommodate various manipulations in task demands and stimulus presentation. In the present study, for example, we adapted the wordlikeness judgment paradigm into a new variant (discussed below) designed to demonstrate sensitivity to *n*gram frequency among preschoolers.

Sensitivity to subword regularity in the preschool setting

To what extent are preschoolers sensitive to subword orthographic regularity? Researchers have addressed this question by examining patterns within preschoolers' expressive spelling (e.g., Apel et al., 2006; Pollo et al., 2009; Treiman et al., 2017). Here, an important distinction is made between receptive and expressive spelling, with the former referring to one's skill in identifying (passively) correctly spelled words and the latter referring to one's skill in correctly spelling (actively) to dictation. Using an expressive spelling task, Apel and colleagues (2006) found that preschoolers' spelling patterns of recently learned words were influenced by orthographic and phonological frequency. In other words, preschoolers were more likely to use frequently occurring letter clusters when spelling words than infrequently occurring letter clusters. Importantly, preschoolers' sensitivity to subword orthographic regularity—seen within their expressive spelling patterns—predicts spelling performance on standardized tests years after initial assessment in preschool (Kessler, Pollo, Treiman, & Cardoso-Martins, 2013). Similar research has shown that preschool children are more likely to use common bigrams in their expressive spelling patterns than uncommon ones (Pollo et al., 2009). Most recently, Treiman and colleagues (2017) observed wordlike patterns of subword orthographic regularity within preschoolers' expressive spellings and linked it conceptually with the emerging statistical learning perspective (Mano, 2016; Pollo, Treiman, & Kessler, 2008; Treiman & Kessler, 2014).

Notwithstanding the importance of the above studies with preschoolers, they have relied on expressive rather than receptive tasks, which is notable because reading is primarily a receptive task. A receptive measure of sensitivity to subword orthographic regularity, such as the wordlikeness judgment task, has the potential to build upon results from expressive tasks by exploring perceptual and task-related factors associated with preschoolers' sensitivity to *n*gram frequency. For example, it is theoretically important to determine whether preschoolers are sensitive to the frequency of particular grain sizes of subword orthography (e.g., unigram, bigram, and/or trigram). On this point, theoretical and empirical work suggests that bigrams play a uniquely important role in orthographic processing (e.g., Chetail, 2017; Dehaene, Cohen, Sigman, & Vinckier, 2005; Vinckier, Qiao, Pallier, Dehaene, & Cohen, 2011; Whitney, 2001). Based on this work, it is conceivable that preschoolers develop sensitivity to the statistical regularity of bigrams before any other grain size (e.g., unigram, trigram).

Present study

The overarching aim of the present study was to establish preschoolers' sensitivity to *n*gram regularity with a receptive measure (i.e., wordlikeness judgment). This research builds upon existing research with expressive spelling and extends the use of the wordlikeness judgment into the preschool setting. Our specific aim was to examine the association between preschooler age and sensitivity to *n*gram regularity and whether such sensitivity (if present) is specific to certain grain sizes (e.g., unigram, bigram, and/or trigram). We utilized correlational and mixed analyses of variance to determine the association between preschooler age and grain sizes of subword orthographic regularity. These efforts allowed us to examine the earliest moments in which receptive sensitivity to subword orthographic regularity may be observed.

To achieve our aim, we used a computerized version of the wordlikeness judgment paradigm. Specifically, we modified the wordlikeness judgment paradigm wherein the task involved presenting pairs of *n*grams individually and in isolation. In this task, attentional focus is constrained to already chunked letter clusters (i.e., unigram, bigram, and trigram; e.g., CH, SP, and NTL) presented in isolation. Results from the present study were intended to fill a critical gap within the literature and advance ongoing theory development, as no study to date has examined receptive sensitivity to *n*gram frequency among preschoolers.

Methods

Participants

Children between ages 4 and 5 years were recruited from a university-based preschool center designed around a constructivist-learning model. This center serves students enrolled through Head Start and tuition. Our sample consisted of 54 children (27 girls) who ranged in age between 49 and 72 months (M_{months} = 60, SD_{months} = 5). Although we did not assess literacy in children individually, preschool-wide teacher ratings from the Literacy scale of the Teaching Strategies GOLD (Heroman, Burts, Berke, & Bickart, 2010) were obtained. Teacher ratings on the Literacy scale (preschool-wide level) for children enrolled through Head Start and children enrolled through tuition were comparable to national normative samples. The Technical Summary of Teaching Strategies GOLD (Lambert, Kim, Taylor, & McGee, 2010) reports reliabilities for all scales ranging from .95 to .98.

Procedure

Children were tested individually in a quiet room. Experimental tasks unrelated to reading were also administered, but those data are not relevant to the present study and are thus not reported. The Independent Review Board approved all procedures. Parents signed the informed consent and children assented prior to data collection.

Subword orthographic task

Children performed a computerized orthographic task (i.e., Word-Making), described as a "game" to the children. Table 1 provides items used in the Word-Making task (full set). The Word-Making game was designed to determine sensitivity to *n*gram regularity as a function of grain size. There were three trial types (see Table 1): pairs of unigrams, pairs of bigrams, and pairs of trigrams. Items were presented in isolation without positional information from a pseudoword context. Items within pairs differed in the frequency with which they appear within the English orthography. Thus, items were either high-frequency or low-frequency units (6 unigram trials, 10 bigram trials, and 10 trigram trials). Stimulus statistics were calculated with MCWord (Medler & Binder, 2005). As such, *n*gram pairs differed in their mean spatially unconstrained frequency. Table 2 displays stimuli statistics for all trial types in terms of unigram, bigram, and trigram structures. Spatially unconstrained frequency statistics differed significantly between high- and low-frequency units ($ps < .05$) such that the high-frequency units always comprised *n*grams that are relatively more frequent within the English orthography.

Trials were presented in blocked order, randomized within trial type, and within left-right position. Only consonant letters were used. Ten unfilled character spaces separated the items.

Table 1. Stimuli used in the word-making task/game.

Condition	High Frequency	Low Frequency
Single letter	t, n, s, h, r, l	z, q, j, x, k, v
Bigrams	vn, tz, tx, qt, pn,sp, th, tr, wh	mh, kt, fn, dt, cm
Trigrams	ntl, nch, sch, tch, ght ldr, rld, mpl, ttl, rds	xcs, wgn, vds, spx, spv, tgh, rrc, qnv, bms, bft

Table 2. Stimulus statistics for the word making task/game.

Mean Positionally Unconstrained	Mean Positionally Unconstrained	Mean Positionally Unconstrained	
Condition	unigram frequency	bigram frequency	trigram frequency
High-frequency unigram	266,712 (72,870)	–	–
Low-frequency unigram	15,308 (16,501)	–	–
High-frequency bigram	208,337 (63,720)	22,291 (36,958)	–
Low-frequency bigram	190,703 (42,304)	12 (11)	–
High-frequency trigram	224,467 (54,756)	9,230 (3,800)	1,573 (1,638)
Low-frequency trigram	156,537 (41,090)	2,129 (1,401)	0(0)

Note. Standard deviations are reported within parentheses. Orthographic statistics are obtained from MCWord (Medler & Binder, 2005), an online orthographic word-form database based on the CELEX efw.cd file. Mean positionally unconstrained frequency refers to the averaged frequency (per million) of the structure (i.e., unigram, bigram, and trigram) within a word regardless of its position or the word length (hence, positionally unconstrained). For example, the "ba" in "bat" is considered the same as the "ba" in "tabasco" (example taken from MCWord; Medler & Binder, 2005).

Preschoolers pointed to which *n*grams would be most helpful in making new words (an examiner sitting adjacent to children recorded their responses). Specifically, children were given the following instructions:

> I want you to help me make up words. I'm going to show you some letters and groups of letters, two at a time. I want you to point to the letters or groups of letters that will be most helpful in making new words. If you are not sure of the answer, make your best guess.

Stimulus presentation and response were controlled with DirectRT. A Dell laptop computer was situated approximately 55 cm in front of the children. All stimuli were presented in lowercase, Times New Roman, white font against a black background. Letters were presented in lowercase font to reflect the style most likely to be encountered within children's books. Preschoolers were not given feedback on their performance.

Data analyses

There were a total of three trial types in the Word-Making task: unigram trials, bigram trials, and trigram trials. The dependent variable was the proportion of high-frequency choices for each trial type. Zero-order, one-tailed correlations between age (months) and trial type performance were calculated to determine which type of subword regularity represents the earliest moment of sensitivity. Here, because three correlations were calculated, a Bonferroni correction was applied to the alpha level to control for Type I error (.05/3 = .016).

In addition to zero-order correlations, we performed a mixed repeated-measures ANOVA for the Word-Making task. Age group was the between-group factor, whereas trial type was the within-group factor. Age group was identified by a medium split (mean $age_{younger}$ = 55 months, *SD* = 2.9 months; mean age_{older} = 65 months, *SD* = 3.0 months). Notably, older preschoolers did not differ from younger preschoolers in terms of educational history, length of school day, and amount of instruction. Trial types for the Word-Making task were unigram trials, bigram trials, and trigram trials. Finally, one-sample *t* tests determined whether the proportion of high-frequency choices differ from chance performance (i.e., 50%) between younger and older preschoolers for each trial type of the Word-Making task.

Results

Table 3 displays means, ranges, and skewness/kurtosis for all trial types of the Word-Making task. Performance for all trial types was normally distributed, with mean accuracy ranging from 46% to 52%. Table 4 displays correlations between performance accuracy and age (months), separately for each trial type. Notably, the only statistically significant correlation observed with age was with bigram trial types, as depicted in Figure 1A. Figure 2 displays nonsignificant associations between age (months) and the unigram and trigram conditions of the Word-Making.

For the Word-Making task, a mixed repeated-measures ANOVA with age group as the between-subject factor (older preschoolers [*n* = 27], younger preschoolers [*n* = 27]) and trial type as the

Table 3. Average proportion of high-frequency choices during the word-making task/game (*N* = 54).

Mean Trial Type	Min + *SD*	Skewness/ Max	Kurtosis
Single letter frequency	.46 ± .22	.16–1.00	.707/-.286
Bigram frequency	.52 ± .18	.10–1.00	−.035/-.075
Trigram frequency	.47 ± .14	.20-.80	−.045/-.261

Note. In the Word-Making task, pairs of orthographic stimuli were used that varied grain size (i.e., single letter, bigram, trigram), with one item of the pair being more frequent than the other (e.g., **t** z; **ch** vn; **ntl** xcs; high-frequency items in bold for illustration).

Table 4. Zero-order correlations between average proportions of high-frequency choices in the word-making task/game and age (N = 54).

Trial type	Age
Single letter frequency	.133
Bigram frequency	.354*
Trigram frequency	.141

Note. Specifically, the alpha level for the correlation between age and bigram frequency (in the Word-Making game) was .004.
*p < .01 (corrected alpha level of .016 following a Bonferroni correction, .05/3).

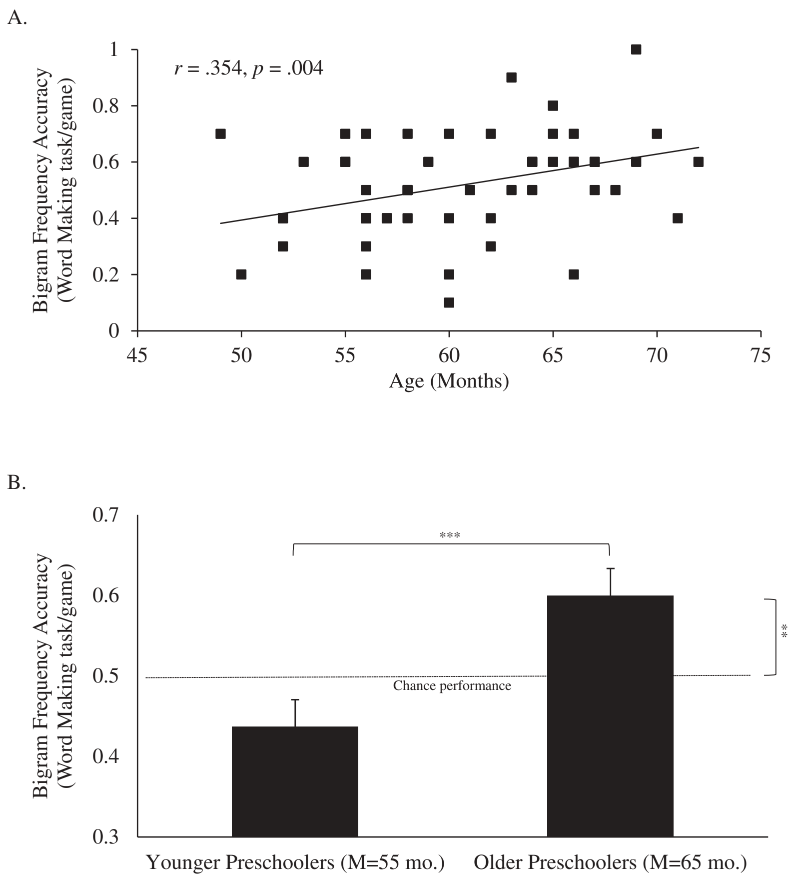

Figure 1. Bigram frequency accuracy in word making task in relation to age (months).

within-subjects factor (unigram frequency, bigram frequency, trigram frequency) showed a non-significant main effect of trial type on word-making judgments, $F(2, 51) = 1.303$, $p = .281$, $\eta_p^2 = .049$. The interactive effect of age group and trial type on Word-Making judgments was significant, $F(2, 51) = 4.132$, $p = .022$, $\eta_p^2 = .139$. Follow-up ANOVAs showed that age group had a significant effect on Word-Making judgments for bigrams, $F(1, 52) = 12.083$, $p = .001$, $\eta_p^2 = .189$, but not for accuracy for unigram and trigram frequency, $F(1, 52) = .245$, $p = .623$, $\eta_p^2 = .005$; $F(1, 52) = .036$, $p = .851$,

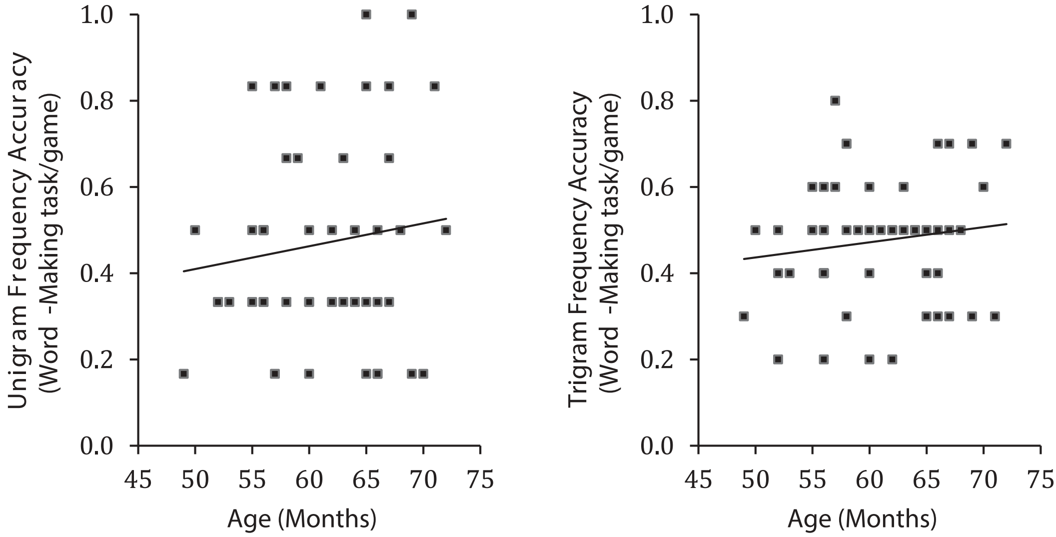

Figure 2. Scatterplots showing nonsignificant associations between age (months) and stimulus conditions of word-making (Unigram, Trigram) task/game. *Note*. $^{**}p \leq .01$. $^{**}p \leq .001$.

$\eta_p^2 = .001$; respectively. Regarding accuracy for bigram frequency, older preschoolers had a significantly higher proportion of high-frequency choices (M = 60%, SD = .17) than younger preschoolers did (M = 43%, SD = .17). As displayed in Figure 1B, among the younger preschoolers, performance on the bigram trials (M = 43%, SD = .17) was not statistically different from chance, $t(26) = -1.885$, $p = .071$. In contrast, among the older preschoolers performance on the bigram trials differed significantly from chance, $t(26) = 3.039$, $p = .005$ (see Figure 1B). Preschoolers, regardless of their age group, did not perform statistically different from chance on the unigram and trigram trial types of the Word-Making task (ps > .05).

Discussion

The present study examined whether preschoolers demonstrate sensitivity to the statistical regularity within subword orthography using a receptive task adapted from the wordlikeness judgment paradigm. Broadly, this receptive task was difficult for the preschoolers. The subtle frequency by which letters and letter combinations appear within print is not an obvious fact for preschoolers. Even so, there was some success with the wordlikeness judgment paradigm within the preschool setting, with results contributing to our understanding of sensitivity to *n*gram frequency in the early years of education. Specifically, we found that older preschoolers showed sensitivity to bigram frequency when the bigram was presented in isolation. Although younger preschoolers performed at chance, there was a significant correlation between bigram sensitivity and age, with older preschoolers performing above chance. Although we did not directly assess print exposure, we reasoned that older preschoolers would have been exposed to greater amounts of print compared to younger preschoolers. This is important because print exposure likely contributed to older preschoolers demonstrating receptive sensitivity to the frequency of bigrams but not younger preschoolers. Older preschoolers are likely to have had more opportunities with print to develop sensitivity to its statistical regularities.

Present results suggest that the earliest moments of observable sensitivity to *n*gram frequency appears to be driven by basic orthographic chunks of letter clusters regardless of their spatial position within words. The size of attentional focus may be an important methodological manipulation needed to demonstrate receptive sensitivity to subword orthographic regularity among preschoolers. Here, it appears that letter patterns need to be perceptually salient and already chunked for preschoolers. Below is a discussion of present findings, future research directions, and practical implications.

Contributions to existing research

To date, relevant research has shown that (1) preschoolers demonstrate sensitivity to subword orthographic regularity within their expressive spelling patterns (e.g., Pollo et al., 2009; Treiman et al., 2017) and (2) children in kindergarten through grade three demonstrate sensitivity to the regularity of subword orthography on a receptive measure (i.e., wordlikeness judgment task; Apel et al., 2013; O'Brien, 2014). Regarding the first point, we build upon findings from expressive spelling research conducted by Pollo and colleagues (2009) in showing that preschoolers demonstrate sensitivity to bigram frequency on a receptive measure of such sensitivity. Regarding the second point, we broaden the applicability of the wordlikeness judgment paradigm in showing that variants of this paradigm may be applied to preschoolers. Within this context, present findings make two theoretical contributions to existing research.

First, present results suggest that the spatial regularity of letter patterns within words is not an important factor in the receptive assessment of early *n*gram frequency sensitivity, at least among preschoolers. Rather, displaying letter clusters in isolation without any spatial context (i.e., pseudoword) appears sufficient to demonstrate the earliest moments of subword orthographic sensitivity, particularly in terms of the wordlikeness judgment paradigm. Notably, the Word-Making task presented *n*grams in isolation (e.g., CH, SP, and NTL); thus, constraining the size of attentional focus to particular letter combinations. This is notable because it is conceivable that the size of attentional focus is an important manipulation for preschoolers whereby *n*grams need to be "already chunked" and made perceptually salient for preschoolers to demonstrate sensitivity to the statistical regularity of subword orthography.

The size of attentional focus, and the role that it plays in the present study, is related to research linking visual attention with reading acquisition (e.g., Franceschini, Gori, Ruffino, Pedrolli, & Facoetti, 2012). For example, Franceschini and colleagues (2012) found that preschoolers' visual spatial attention (measured with a serial search and spatial cueing task) predicted reading performance in Grades 1 and 2, even after controlling for nonverbal IQ, speech-sound processing, and nonalphabetic cross-model mapping. In fact, Franceschini et al. (2012) reported that 60% of poor readers in Grades 1 and 2 displayed visual spatial attentional problems when they were preschoolers. Moreover, Mano (2016) highlights the importance of (visual) attentional control in dynamically interacting with incidental statistical learning, which underlies developing sensitivity to the statistical regularities of subword orthography. Related to the issue of visual attentional control is evidence suggesting that relatively well-established word reading skills lead to increases in visual perceptual span (Sperlich, Meixner, & Laubrock, 2016). These studies raise a potential limitation: The width of the attentional focus of preschoolers may have been too narrow to be able to detect orthographic regularity within larger orthographic units such as trigrams. Preschool children may not have sufficiently large orthographic representations on which to inform wordlikeness judgments with trigrams.

Second, present findings suggest that subword structures have to be bigrams, not trigrams or unigrams, to elicit the earliest moments of *n*gram frequency sensitivity—at least on a receptive measure of such sensitivity. Grain size appears to be an important parameter of the wordlikeness judgment paradigm in assessing sensitivity to subword regularity among preschoolers. Why bigrams and not unigrams or trigrams? Although we did not have specific hypotheses with respect to grain size, prior research shows that preschoolers' expressive spelling patterns exhibit sensitivity to unigram and bigram frequencies (Pollo et al., 2009). Yet we found that early sensitivity was limited to bigrams, not unigrams (or trigrams for that matter). It is conceivable that differences between expressive (e.g., spelling) and receptive (e.g., wordlikeness judgment) task demands explain why only bigram sensitivity was observed in the present study. For example, observing sensitivity to unigram frequency may be dependent on expressive assessments such as spelling. It may be easier for preschoolers to retrieve frequently occurring single letters in their writing, whereas the relative frequency of single letters may be too subtle for preschoolers to perceive on a receptive measure. A lack of receptive frequency effects for unigrams among preschoolers might stem from the fact that single letters are presented equally often in the context of alphabetic learning. Moreover, preschoolers were exposed to alphabet learning strategies that may explain in part why frequency sensitivity for unigrams was not observed.

Bigrams may have a special status in reading. In fact, as alluded to above, some mathematical and neurobiological models of word reading place unique emphasis on bigram structures (Dehaene et al., 2005; Whitney, 2001). A common assumption within these models is that bigram processing represents a midpoint between the processing of single letters and words. For words comprising four or more letters, visual word recognition may not be achieved solely by combining single letters into sequences. Instead, visual recognition of such words may be achieved by processing constituent bigrams. Present results suggest that bigrams represent an initial level wherein perceptual tuning to the statistical regularity of *n*grams occurs. It is conceivable that older students (i.e., ages 7–9) demonstrate sensitivity to the statistical regularities of *n*grams other than bigrams, such as trigrams (Conrad et al., 2013). Future longitudinal research is needed to determine whether preschoolers who demonstrate sensitivity to the statistical regularity of bigrams go on to later demonstrate sensitivity to the statistical regularities of trigrams in later school years.

Limitations

There are several limitations of this study that are worth mentioning. First, some of the high-frequency items in the Word-Making task were common graphemes (e.g., CH, SH), which may confound present results insofar as some preschoolers may have been exposed to letter-sound mappings. Second, we did not assess each child's alphabetic knowledge, which is a notable limitation given that it is important to determine whether sensitivity to the statistical frequency of bigrams is associated with letter-sound mappings or if such sensitivity is observed independent of alphabetic knowledge.

Though *n*grams in the Word-Making game were not embedded within larger orthographic units such as pseudowords, position effects may still have contributed to performance on this game. The position of first and second letters of trigrams, for example, may have contributed incidentally to judgments in the Word-Making game. Related to this point is the fact that many of the trigram pairs in the Word-Making game were not matched on constituent bigram frequency. For example, the bigram frequency of some high-frequency trigram items was relatively higher than the bigram frequency of some low-frequency trigram items. However, if bigram frequency influenced performance in the bigram condition of the Word-Making game, then it should have also influenced performance in the trigram condition of the same game, but it did not. The fact that differences in bigram frequency did not influence performance in the trigram condition highlights the special status of bigrams presented in isolation while demonstrating insensitivity to the frequency of subword orthographic structures larger than bigrams (e.g., trigram).

Finally, it is important to note that present findings speak to the assessment of *n*gram frequency sensitivity and not the development of such sensitivity. These limitations, however, do not subtract from the basic observation made in the present study, which is that preschoolers demonstrate sensitivity to the statistical regularity of bigrams when presented in isolation and without positional information.

Future directions and practical implications

Future research needs to examine whether performance on the present Word-Making task is associated with alphabetic knowledge, word reading, verbal reasoning, oral language measures, and/or print exposure. It would be interesting if preschoolers who are not yet able to sound out words or to produce phonologically plausible renditions of words in spelling have some knowledge about common letter groupings. Such observations would further highlight sensitivity to subword orthographic regularity as an important facet of literacy acquisition.

Future research is also needed to determine if children are capable of articulating the reason for their wordlikeness judgments. Are children basing their wordlikeness judgments on the statistical structure of letter combinations or perhaps on phonological cues? The articulated reasons for wordlikeness judgments may reveal insights into the cognitive processes associated with *n*gram frequency sensitivity. It is important to reiterate here the present study speaks to the assessment of *n*gram frequency sensitivity and not to its development. As such, future research is needed to determine whether variables identified in the present study as being important for assessment (i.e., perceptual saliency, bigrams) are also important for developing sensitivity to the frequency of subword orthography. These sorts of studies, if asked with more background knowledge about participants (e.g., reading ability, print exposure, home literacy environment), may reveal new insights into early orthographic processing and its relation to the development of basic reading skills.

Present results—when combined with existing research linking sensitivity to subword orthographic regularity to reading outcomes (e.g., Cassar & Treiman, 1997; Conrad et al., 2013; O'Brien, 2014; Pacton et al., 2001; Rothe et al., 2014; Stanovich & Siegel, 1994)—have practical implications for reading instruction and remediation. First, it is conceivable that developing sensitivity to the statistical regularity of letter combinations within the English orthography represents a facet of typical reading development. Indeed, present results show that children as young as preschool age appear capable of demonstrating sensitivity (albeit subtly) to the positionally unconstrained frequency of bigram units, using a receptive task (i.e., wordlikeness judgment). It is unlikely that this sensitivity emerged from explicit reading instruction (i.e., children were not instructed to compute bigram frequencies), which suggests that a type of incidental learning mechanism contributed to the development of such sensitivity. In fact, according to recent theoretical work (Mano, 2016), sensitivity to the statistical regularity of *n*grams emerges through interplay between incidental statistical learning and attentional control. If true, then educators may consider developing novel instructional methods for students to attend to reoccurring letter patterns in order facilitate statistical learning.

A second practical implication involves reading remediation. Although more speculative in nature, it is conceivable that some variance in reading difficulty (and/or reading disability) is accounted for by insensitivity to the statistical regularities of letter combinations. According to recent theoretical work (Mano, 2016), insensitivity to the statistical regularities of letter combinations may result from (1) impaired or delayed incidental learning mechanisms, (2) impaired and/or delayed attentional control, or (3) lack of sufficient interplay between incidental learning and attentional control. Indeed, this is very much speculative, but present results help to confirm the special status of frequency-based chunking in early literacy development, and as such, it is conceivable that future forms of reading remediation will involve creating new contexts in which readers can develop sensitivity to frequency-based letter chunks. Clearly, additional research is warranted.

Disclosure statement

No potential conflict of interest was reported by the authors.

Funding

This work was supported by the National Science Foundation (SMA-1263142).

References

Apel, K., Brimo, D., Wilson-Fowler, E. B., Vorstius, C., & Radach, R. (2013). Children develop initial orthographic knowledge during storybook reading. *Scientific Studies of Reading*, *17*(4), 286–302. doi:10.1080/10888438.2012.692742

Apel, K., Wolter, J. A., & Materson, J. (2006). Effects of phonotactic and orthographic probabilities during fast mapping on 5-year-olds' learning to spell. *Developmental Neuropsychology*, *29*, 21–42. doi:10.1207/s15326942dn2901_3

Arciuli, J., & Simpson, I. C. (2012). Statistical learning is related to reading ability in children and adults. *Cognitive Science*, *36*(2), 286–304. doi:10.1111/j.155106709.2011.01200.x

Cassar, M., & Treiman, R. (1997). The beginnings of orthographic knowledge: Children's knowledge of double letters in words. *Journal of Educational Psychology*, *89*(4), 631–644. doi:10.1037/0022-0663.89.4.631

Chetail, F. (2017). What do we do with what we learn? Statistical learning of orthographic regularities impacts written word processing. *Cognition*, *163*, 103–120. doi:10.1016/j.cognition.2017.02.015

Conrad, N. J., Harris, N., & Williams, J. (2013). Individual differences in children's literacy development: The contribution of orthographic knowledge. *Reading and Writing: An International Journal*, *26*(8), 1223–1239. doi:10.1007/s11145-012-9415-2

Dehaene, S., Cohen, L., Sigman, M., & Vinckier, F. (2005). The neural code for written words: A proposal. *Trends in Cognitive Sciences*, *9*, 335–341. doi:10.1016/j.tics.2005.05.004

Franceschini, S., Gori, S., Ruffino, M., Pedrolli, K., & Facoetti, A. (2012). A causal link between visual spatial attention and reading acquisition. *Current Biology*, *22*(9), 814–819. doi:10.1016/j.cub.2012.03.013

Heroman, C., Burts, D. C., Berke, K., & Bickart, T. S. (2010). *The creative curriculum for preschool—Volume 5, Objectives for development & learning: Birth through kindergarten*. Washington, DC: Teaching Strategies, LLC.

Ise, E., Arnoldi, C. J., & Schulte-Körne, G. (2014). Development of orthographic knowledge in German-speaking children: A 2-year longitudinal study. *Journal of Research in Reading*, *37*(3), 233–249. doi:10.1111/j.1467-9817.2012.01535.x

Kessler, B. (2009). Statistical learning mechanisms of conditional orthogrpahic correspondences. *Writing Systems Research*, *1*(1), 19–34. doi:10.1093/wsr/wsp004

Kessler, B., Pollo, T. C., Treiman, R., & Cardoso-Martins, C. (2013). Frequency analyses of prephonological spellings as predictors of success in conventional spelling. *Journal of Learning Disabilities*, *46*(3), 252–259. doi:10.1177/0022219412449440

Kirkham, N. Z., Slemmer, J. A., & Johnson, S. P. (2002). Visual statistical learning in infancy: Evidence for a domain-general learning mechanism. *Cognition*, *83*(2), B35–B42. doi:10.1016/S0010-0277(02)00004-5

Lambert, R. G., Kim, D. H., Taylor, H., & McGee, J. (2010). *Technical manual for the Teaching Strategies GOLD® assessment system*. Charlotte, NC: Center for Educational Measurement and Evaluation.

Mano, Q. R. (2016). Developing sensitivity to subword combinatorial orthographic regularity (SCORe): A two-process framework. *Scientific Studies of Reading*, *20*(3), 231–247. doi:10.1080/10888438.2016.1141210

Mano, Q. R., & Guerin, J. M. (2018). Direct and indirect effects of print exposure on silent reading fluency. *Reading and Writing*, *31*(2), 483–502. doi:10.1007/s11145-017-9794-5

Mano, Q. R., & Medler, D. A. (2017). *Descriptive analysis of subword orthographic structures within the English orthography*. Retrieved from http://homepages.uc.edu/~manoqo/LCAN/top30subwords.html

Medler, D. A., & Binder, J. R. (2005). *MCWord: An on-line orthographic database of the English language*. Retrieved from http://www.neuro.mcw.edu/mcword/

O'Brien, B. A. (2014). The development of sensitivity to sublexical orthographic constraints: An investigation of positional frequency and consistency using a wordlikeness choice task. *Reading Psychology*, *35*(4), 285–311. doi:10.1080/02702711.2012.724042

Pacton, S., Perruchet, P., Fayol, M., & Cleeremans, A. (2001). Implicit learning out of the lab: The case of orthographic regularities. *Journal of Experimental Psychology: General*, *130*(3), 401–426. doi:10.1037/0096-3445.130.3.401

Pollo, T. C., Kessler, B., & Treiman, R. (2008). Three perspectives on spelling development. In E. J. Griogorenko & A. Naples (Eds.), *Single-word reading: Cognitive, behavioral, and biological perspectives* (pp. 175–189). Mahwah, NJ: Erlbaum.

Pollo, T. C., Kessler, B., & Treiman, R. (2009). Statistical patterns in children's early writing. *Journal of Experimental Child Psychology*, *104*(4), 410–426. doi:10.1016/j.jecp.2009.07.003

Rothe, J., Schulte-Körne, G., & Ise, E. (2014). Does sensitivity to orthographic regularities influence reading and spelling acquisition? A 1-year prospective study. *Reading and Writing: An International Journal*, *27*(7), 1141–1161. doi:10.1007/s11145-013-9479-7

Saffran, J. R., Aslin, R. N., & Newport, E. L. (1996). Statistical learning by 8-month-old infants. *Science*, *274*, 1926–1928.

Saffran, J. R., Johnson, E. K., Aslin, R. N., & Newport, E. L. (1999). Statistical learning of tone sequences by human infants and adults. *Cognition*, *70*(1), 27–52. doi:10.1016/S0010-0277(98)00075-4

Sperlich, A., Meixner, J., & Laubrock, J. (2016). Development of the perceptual span in reading: A longitudinal study. *Journal of Experimental Child Psychology*, *146*, 181–201. doi:10.1016/j.jecp.2016.02.007

Stanovich, K. E., & Siegel, L. S. (1994). Phenotypic performance profile of children with reading disabilities: A regression-based test of the phonological-core variable-difference model. *Journal of Educational Psychology*, *86*(1), 24–53. doi:10.1037/0022-0663.86.1.24

Treiman, R., & Kessler, B. (2014). *How children learn to write words*. New York, NY: Oxford University Press.

Treiman, R., Kessler, B., Boland, K., Clocksin, H., & Chen, Z. (2017). Statistical learning and spelling: Older prephonological spellers produce more wordlike spellings than younger prephonological spellers. *Child Development*. doi:10.1111/cdev.12893

Turk-Browne, N. B., Jungé, J. A., & Scholl, B. J. (2005). The automaticity of visual statistical learning. *Journal of Experimental Psychology: General*, *134*(4), 552–564. doi:10.1037/0096-3445.134.4.552

Vinckier, F., Qiao, E., Pallier, C., Dehaene, S., & Cohen, L. (2011). The impact of letter spacing on reading: A test of the bigram coding hypothesis. *Journal of Vision*, *11*(8). doi:10.1167/11.6.8

Whitney, C. (2001). How the brrain encodes the order of letters in a printed word: The SERIOL model and selective literature review. *Psychonomic Bulletin & Review*, *8*(2), 221–243. doi:10.3758/BF03196158

Wright, D.-M., & Ehri, L. C. (2007). Beginners remember orthography when they learn to read words: The case of doubled letters. *Applied Psycholinguistics*, *28*(1), 115–133. doi:10.1017/S0142716406070068