

Statistical and explicit learning of graphotactic patterns with no phonological counterpart: Evidence from an artificial lexicon study with 6–7-year-olds and adults

Daniela Singh<sup>1</sup>, Elizabeth Wonnacott<sup>2</sup>, & Anna Samara<sup>3</sup>

<sup>1</sup>Division of Psychology and Language sciences, University College London, United Kingdom

<sup>2</sup>Department of Education and St Johns College, University of Oxford, United Kingdom

<sup>3</sup>School of Human Sciences, University of Greenwich, United Kingdom

Email: [daniela.singh@education.ox.ac.uk](mailto:daniela.singh@education.ox.ac.uk)

**Abstract**

Children are powerful statistical spellers, showing sensitivity to untaught orthographic patterns. They can also learn novel written patterns with phonological counterparts via statistical learning processes, akin to those established for spoken language acquisition. It is unclear whether children can learn written (graphotactic) patterns which are unconfounded from correlated phonotactics. We address this question by inducing novel graphotactic learning under incidental versus explicit conditions. Across three artificial lexicon experiments, we exposed children and adults to letter strings ending either in singlets or doublets (that share the same pronunciation, e.g., *s* vs. *ss*) depending on the preceding vowel. In post-tests, children and adults incidentally generalized over such context-based constraints that varied in complexity. Explicit instruction further benefitted pattern generalization, supporting the practice of teaching spelling patterns, and there was a relationship between explicit learning and literacy scores. We are first to demonstrate that statistical learning processes underlie graphotactic generalizations among developing spellers.

**Keywords:** statistical learning, explicit instruction, graphotactics, Bayes factors

**Statistical and explicit learning of graphotactic patterns with no phonological counterpart:  
Evidence from an artificial lexicon study with 6–7-year-olds and adults**

Written language is a highly patterned domain, and as such, visually presented words comply with regularities and constraints that are well explained in statistical terms. For example, English words never begin with *ck*: Simple, deterministic patterns of this sort are easy to verbalize and overtly taught in school. Other patterns (e.g., “consonants often double after single-vowel rather than double-vowel spellings”) receive less attention in formal literacy instruction at least in part due to their complex, probabilistic nature (Kessler, 2009).

Patterns can be broadly categorized as follows: Firstly, they may directly parallel spoken language (phonotactic) restrictions (e.g., in English, words do not begin with /ŋ/ and accordingly, written words cannot have *ng* in beginnings), or, secondly, they can be orthographic but reflect the influence of phonetic environment (e.g., /ε/ is particularly likely to be spelled as *ea* when the word ends in /d/; e.g., *dead*). Lastly, patterns can be purely visual (graphotactic) in nature with no phonological counterpart (e.g., *dd* does not begin written words, as in *\*ddoll*, /d/ does).

There is experimental evidence that spellers gradually pick up on these patterns over the course of their development and use all three types by the time they are adults (Treiman, 2018). Artificial lexicon studies have begun to explore whether this ability is underpinned by *statistical learning* processes, akin to those thought to play a role in spoken language acquisition; however, no learning experiment has, to date, conclusively demonstrated the learning of purely visual graphotactic spelling patterns in children. Are statistical learning processes only implicated in literacy acquisition when patterns operate across spoken and written language (e.g. words do not begin with /ŋ/ or *ng*)? The current study investigates, in three learning experiments, young children’s and adults’ ability to learn written (purely graphotactic) constraints controlling for the possibility that learning could be underpinned by processes operating over spoken language. We

also address the question of whether learning in the experiments occurs without awareness (incidental learning) and whether nevertheless, explicit awareness provided through instruction is beneficial. Finally, we explore how learning spelling patterns in the laboratory relates to individual differences in children's literacy skills.

### **Learning spelling constraints from natural language**

As children learn to read and spell, they increasingly develop tacit knowledge of statistical properties of printed and spoken words (Pollo, Treiman, & Kessler, 2008). This view of literacy development has been extensively explored in relation to the ability to pick up on patterns that regulate when and where consonants double. In English, for example, doublets only appear within word middles (e.g., *bunny*), or word endings (e.g., *bell*), and rarely occur for letters such as *k* and *y*.

Frequency-based influences of this sort are sometimes seen in naturalistic spellings of preliterate children (Treiman, 1993) and emerge in carefully designed experiments. Cassar and Treiman (1997) showed, for example, that English-speaking children (5–11-year-olds) and adults prefer permissible nonwords (e.g., *baff*, *yill*, *geed*) to those that violate some of the written constraints described above (e.g., *\*bbaf*, *\*yyil*, *\*gaad*), and become more accurate and sensitive to more complex influences as their print exposure increases with age (see also Treiman, Kessler, Boland, Clocksin, & Chen, 2017). Importantly, these effects are not an artifact of sensitivity to overall (singled) letter frequencies (e.g., if *yill* may have been preferred over *\*yihh* because *l* is more frequent than *h*). This was shown by Pacton, Perruchet, Fayol and Cleeremans (2001) in a wordlikeness task, in which participants were asked to choose which of two nonwords more like a real word. From 6 years of age, French-speaking children preferred the nonwords whose critical consonants were more frequent not only in single form but also in double form.

The patterns shown to affect performance in the experiments above also influence child and adult participants' ability to learn and recall novel spellings. This was shown in more recent work

by Pacton and colleagues (Pacton, Sobaco, Fayol, & Treiman, 2013) who presented 9-year-old French speakers with legally and illegally spelled French nonwords embedded within stories to be read silently for meaning. They found that words that violated double patterns in word beginning and medial positions (Experiment 3) were remembered poorly and illegally spelled words were regularized, mainly via omission errors (i.e., dropping one consonant of a doublet). Qualitatively similar findings have been also seen with French-speaking adults who were incidentally exposed to the nonwords within texts or in isolation (Sobaco, Treiman, Peereman, Borchardt, & Pacton, 2015).

Critically, the studies discussed above each explored unconditional spelling regularities. These may help learners resolve some spelling irregularities, but extensive analyses of the English orthography show that they are less powerful cues than patterns that condition spellings based on surrounding context (Kessler & Treiman, 2001; Treiman & Boland, 2017; see also Treiman & Kessler, 2019 and Venezky, 1970 for context effects in the reading direction). Medial vowels, for example, influence following consonants: short vowels (for American English, /æ/, /e/, /ɪ/, /ɑ/, /ʌ/, /ʊ/) are more likely to be followed by consonantal doublets than singlets (e.g., *supper*) and long vowels tend to be followed by singlets (e.g., *super*). Can children also exploit various conditional spelling regularities?

One type of conditional pattern, namely, phonological context influences on consonant doubling, were assessed in one condition of Cassar and Treiman (1997) by presenting (visually and auditorily) participants with nonwords featuring medial single/double consonants preceded by a short vowel (e.g., /'tɛbɪf/; “is it *tebif* or *tebbif*?”) or a long vowel (e.g. /'sobæp/; “is it *sobap* or *sobbap*”). The English-speaking adults and 11–12-year-olds (but not the young children) were above chance at choosing items embedding correct long (i.e., *sobap* for /'sobæp/) and short (i.e., *tebbif* for /'tɛbɪf/) vowel transcriptions and this was taken to suggest that phonological context affects children’s preferences later in development relative to constraints on positions and letter identity.

The phonological pattern on vowel pronunciation described above, however, is not the only possible cue to consonantal doubling. Hayes, Treiman, and Kessler (2006) investigated sensitivity to a *graphotactic* cue that operates independently from phonology: doublets occur more often after a single- than double-letter spelling (e.g., *Jeff* vs. *deaf*; *bedding* vs. *heading*). Similar to Cassar and Treiman (1997), they asked English-speaking children to choose the most word-like item between two nonwords that either conformed to or violated this pattern (e.g., *vaff* vs. *\*vaf*; *vaiif* vs. *\*vaiiff*) and also used a nonword production task. Pattern-conforming performance was shown in both, but the nonword production results are particularly important in one additional way: The graphotactic influence can be hard to distinguish from the phonological pattern on vowel pronunciation because, in English, short vowels almost always take single-letter spellings (e.g., *tell*) while long vowels often take two-letter spellings (e.g., *tail*). In many instances, therefore, the phonological and graphotactic cues correlate with each other. In the 7–8-years-olds' spelling attempts, though, doubling was less likely to occur after two-vowel letters even when they represented (somewhat unconventionally) a short vowel. For example, if a short vowel /æ/ in /sæf/ was spelled as *ae*, *f* was more likely to be produced over *ff*. Thus, the graphotactic pattern influenced children's performance above and beyond the phonology. This has been since also shown with adults using both monosyllabic and disyllabic stimuli (Treiman & Kessler, 2015; Treiman & Boland, 2017; Treiman & Wolter, 2018).

To summarize, skilled (adult) and developing readers are sensitive to a range of spelling patterns that can guide their choice in ambiguous spelling situations. These effects are extensively demonstrated among English-speaking and French-speaking children—whose orthographies embed highly inconsistent relationships between graphemes and phonemes—but are also seen among children learning a more consistent orthography (e.g., Spanish-speakers: Carrillo & Alegria, 2014; Finnish-speakers: Lehtonen & Bryant, 2005). Questions of learnability, however, are hard to address in studies that manipulate children's (and adults') sensitivity to patterns in their actual

orthography. For example, the studies above cannot strictly establish the exact age by which different patterns are acquired and, relatedly, how many reading exposures are necessary for pattern sensitivity to emerge. Answers to these questions are important in order to evaluate theoretical claims regarding when children master different spelling milestones. We use artificial language learning methods to investigate how early graphotactic knowledge emerges and evaluate our findings against the predictions of stage-based models of spelling development (e.g., Frith, 1985; Gentry, 1982)

### **Learning graphotactics in artificial lexicons**

Artificial learning experiments are well-established methods for exploring the types of regularities infant and adult learners can acquire in different domains of knowledge. These experiments originate from the seminal work of Saffran and colleagues (Saffran, Aslin, & Newport, 1996) and involve incidental exposure to patterns and regularities that are made by the experimenter (i.e., that are novel in nature), making participants' previous knowledge irrelevant. In a typically brief (a single session or few sessions long) experiment, participants are trained with pattern-embedding stimuli (an auditory stream of syllables, letter strings, sequences of shapes) and are subsequently asked to judge the legality of trained and (more critically) generalization stimuli that conform to or violate the patterns. This method has been used to demonstrate a role of statistical language learning in various domains: word segmentation (Saffran et al., 1996), syntax (Mintz, 2003; Reeder, Newport, & Aslin, 2013; Wonnacott, Newport, & Tanenhaus, 2008) and—most similar to the current work—learning of *phonotactic constraints*: Learning these constraints on legal and illegal sound distributions is key for learning L1 phonology, and has been demonstrated with novel (artificial) speech sound distributions in infant and adult experiments (Chambers, Onishi, & Fisher, 2003; Onishi, Chambers, & Fisher, 2002).

Importantly, the experiments above have all used *spoken* language, and the authors have made links with processes involved in child first language acquisition. Samara and Caravolas (2014) were the first to extend this methodology to study literacy development. They created a child-appropriate task whereby 7-year-olds and adults saw nonwords and were told that they were going to play games with words from an alien language. There was no mention of any patterns or rules; however, all of the stimuli embedded novel patterns. Two types of pattern were investigated across two experiments – *positional constraints*, e.g., “*t* only begins words” “*f* only ends words” as well as *contextual constraints* where consonant usage was conditioned on the middle vowel, e.g. “*t* can be followed by *o*, never *e*” as well as “*f* can be preceded by *o*, never *e*”. Generalization was tested, in the same session, using legality judgments: Participants were shown novel stimuli that either conformed to or violated the trained patterns and were asked to decide if these could be part of the language they were exposed to. Both children and adults were able to discriminate between legal and illegal stimuli (e.g., for the positional pattern accepting *tof* as conforming to the pattern but rejecting *fof* as not conforming; for the contextual pattern accepting *tof* as conforming and rejecting *tef* as not conforming), although for both participants performance was much stronger for the simpler positional constraints (e.g., for children, 64% vs. 52%, respectively). Samara, Singh, and Wonnacott (2019) replicated and extended the finding among English- and Turkish-speaking children both of whom generalized over constraints on vowel and consonant co-occurrences in either onset or rime positions.

Evidence from studies as those described above demonstrates that statistical learning processes can be applied to the learning of spelling patterns. This supports a view in which this learning plays a role in literacy development, and, at least to some extent, underpins the learning of spelling patterns. However, there is an important limitation in these studies: although learning effects have been interpreted as graphotactic in nature, the constraints had a phonological counterpart, i.e., phonotactic constraints such as, “/t/ can only follow /o/, never /ε/”, alongside

graphotactics (“*t* can only follow *o*, never *e*”). It is, thus, possible that *phonotactic sensitivity* of the type operating in spoken language contributed to learning effects. To the best of our knowledge, only two studies have controlled for this confound but the experiments either featured adults or very simple (positional) constraints.

Chetail (2017) used stimuli comprised of characters unknown to the subjects (letters of the Phoenician script) to investigate adults’ learning over the distribution of co-occurring letters, in familiar or novel positions within the stimuli. She found above-chance discrimination between patterned and random character sequences in a wordlikeness task (whereby participants were asked to judge which of two stimuli was more like the words seen in a previous exposure phase) and evidence that frequency-based learning is represented in the visual word recognition system as seen by participants’ faster reaction times for high-frequency relative to low-frequency characters in a speeded detection task. Given that participants were advanced (adult) spellers, though, it is not possible to draw conclusions regarding statistical learning in children—the population of interest for literacy acquisition.

Nigro, Jiménez-Fernández, Simpson and Defior (2016) made a step in this direction by manipulating unconditional constraints on positions in a study with 8-year-old Spanish-speaking children with and without dyslexia: In one of two learning experiments, children were incidentally exposed to four-character-long nonlinguistic stimuli (sequences of shapes) which adhered to novel constraints, such as, “stimuli only start with *Shape1*, *Shape3*, *Shape5*, *Shape7*—never with *Shape2*, *Shape4*, *Shape6*, *Shape8*”. A legality test involving judging novel stimuli’s well-formedness showed that the constraints were reliably learned by typically developing children and, to some extent, by children diagnosed with dyslexia. Thus, this study established that children can learn visually based constraints which have no phonological counterparts. However, there is an important caveat: The constraints tested were all positional—which can be, as noted above, considerably easier for children to learn in artificial lexicon experiments relative to contextual constraints

(Samara et al., 2014). Children's early invented spellings often conform to unconditional patterns of this sort (e.g. children rarely violate the positional rule "words cannot begin with *ck*") (Treiman, 1993), however, the majority of experimental and longitudinal work suggest that the influence of conditional patterns emerges later (Cassar & Treiman, 1997; Caravolas, Kessler, Hulme, & Snowling, 2005; Treiman & Kessler, 2019). Establishing when conditional patterns are learned is important because these are not only harder for children but also (as noted above) more powerful in predicting spellings, at least for the English orthography (Kessler & Treiman, 2001; Venezky, 1970). The present study asks whether conditional context-based graphotactic constraints with no phonotactic counterpart can be learned by children (and adults) under brief incidental conditions.

### **Relationships between statistical learning and measures of literacy**

If children's ability to extract statistical regularities from carefully designed stimuli taps into the process of learning real spelling patterns, one would expect positive associations between their lab-based performance and their (real life) literacy skills. Demonstrating this link would serve to validate statistical learning as methods with relevance for literacy but would be also theoretically important: Much variance in children's spelling skills remains unexplained once the influence of three well-established predictors (phoneme awareness, letter knowledge, and rapid automatized naming) has been accounted for (Caravolas et al., 2012), raising the question of whether statistical learning ability is another potential predictor. Are better learners in statistical learning tasks with artificial orthography better spellers? If so, well-controlled longitudinal research may use these measures to draw causal links between statistical learning and literacy. To the best of our knowledge, only one study with English-speaking 5–6-year-olds has partially addressed this question. Caravolas and colleagues (2005) showed a predictive influence of unconditional sound–letter correspondence knowledge (probabilities with which particular vowel graphemes occur in English words) on spelling performance but no effect of conditional vowel–coda (context-based)

patterns. Here, as a starting point, we investigate concurrent correlations between lab-based graphotactic learning performance under implicit vs. explicit conditions and performance on literacy measures.

The broader relationship between statistical learning and measures of literacy has received much recent attention (e.g., Frost, Siegelman, Narkiss, & Afek, 2013; Spencer, Kaschak, Jones, & Lonigan, 2015) but hitherto results are inconclusive. Arciuli and Simpson (2012) first reported a positive correlation between reading ability (WRAT; Wilkinson & Robertson 2006) and performance on a visual statistical learning paradigm in children (5- to 12-year-olds) and adults; however, other work has not replicated this finding (Nigro, Jiménez-Fernández, Simpson, & Defior, 2015; Schmalz, Moll, Mulatti, & Schulte-Körne, 2019; West, Vadillo, Shanks, & Hulme, 2018). There are many possible reasons for this discrepancy, but at least in part, null findings may be an artifact of poor statistical properties (e.g., low internal consistency and reliability) of the statistical learning measures (Krishnan, & Watkins, 2019; Siegelman, Bogaerts, & Frost, 2017). Siegelman et al. (2017) argued that most lab-based learning experiments are designed to capture group-level effects, thus, are, by design, unlikely to capture patterns of association: They feature few trials, the level of difficulty is intentionally restricted across test items, and they often suffer from floor effects (i.e., many participants performing at chance levels). In line with this idea, West et al. (2018) reported low split-half test-retest reliabilities (ranging from 0 to .24) for three measures of implicit learning (as opposed to the good reliability in explicit versions of the same tasks) and found no evidence for a relationship between performance in these and literacy and language attainment.

Inconsistent patterns of correlations are not only found with domain-general tasks as those discussed above, but also with measures of children's naturalistic orthographic sensitivity. Treiman and Boland (2017), for example, found that good spellers (according to their WRAT spelling performance) were more likely to double medial consonants in a nonword spelling task measuring graphotactic sensitivity, which, by one interpretation, may be due to their preference for more

frequent letter strings in their orthography. Ise et al. (2014), on the other hand, found no concurrent or longitudinal relationship between their measure of orthographic knowledge (nonword choice task) and reading and spelling ability in German-speaking children.

In sum, it is hard to establish whether statistical property confounds may have concealed a relationship between statistical learning and literacy development, or whether the null findings suggest that there is no fundamental link between variations in the ability to detect statistical structure and literacy acquisition. Findings in this literature may be also distorted by publication biases and false positive (type I) results, both of which are pertinent concerns for psychology and other social sciences (Rosenthal, 1979; Camerer et al., 2018). Finally, when null findings *are* reported, they should be treated with caution for an additional reason: In the frequentist analyses that are typically applied, a nonsignificant result may be evidence for the null or no evidence for any conclusion at all (or indeed evidence against the null). Yet people routinely take a nonsignificant result to indicate that they should reduce their confidence in a theory that predicts a difference. Bayes factors, on the other hand, can provide quantifying evidence of  $H_0$  (Dienes, 2008, 2015). To date, no correlational work has utilized this method of analysis to establish whether nonsignificant associations between statistical learning and literacy skill are substantial evidence against theoretical accounts that predict a relationship or an artefact of data insensitivity.

### **Explicit learning in spelling**

While the evidence reviewed above suggested that aspects of spelling can be acquired without deliberate effort and that knowledge acquired in this manner may be unavailable to retrospection (or “awareness”; Frensch & Rüniger, 2003), spelling patterns and rules are also taught in schools (e.g., “*i* before *e*, except after *c*”; “change *y* to *i* when adding suffix endings”; “*z*, never *s*, spells /z/ at the beginning of a base word”). This is a key difference between spoken and written language acquisition, the former being largely assumed to be learned through incidental exposure

rather than effortfully, that is, through explicit instruction. Given that explicit processes are implicated in learning aspects of written language, is this the optimal method for learning purely graphotactic patterns? If so, is this true across graphotactic patterns that are easy and hard to articulate? Both of these questions are theoretically important and practically relevant for current spelling instruction practices.

Some insights into this question come from within the statistical learning tradition that compared the degree of learning exhibited by participants who were intentionally searching for patterns relative to participants trained under standard (incidental) conditions. These have, however, yielded conflicting results. Arciuli, Torkilsen, Stevens, and Simpson (2014) for example, found nonsignificant differences in adults' incidental and explicit visual statistical learning ability under conditions of short stimulus presentation, whereas Kachergis, Yu, and Shiffrien (2010) demonstrated an explicit condition advantage in a within-subjects design: Participants were better when they were explicitly instructed to count co-occurrence statistics obtained in a cross-situational learning task relative to when they performed the task incidentally.

Turning to orthographic pattern learning, results have been also mixed (Bosman, van Hell, & Verhoeven, 2006; Butyniec-Thomas & Woloshyn, 1997; de Bree, Greelgoed, & van den Boer, 2018; Kemper, Verhoeven, & Bosman 2012; Nunes, Bryant, & Olsson, 2003). Sobaco et al. (2015), for example, directly compared their participants' ability to learn nonwords embedding graphotactic patterns that were either legal or illegal in French, under implicit conditions (whereby the instructions were to simply read the words aloud) and explicit conditions (whereby the nonword spellings were to be memorized) and found better learning recall across spellings in the latter group of participants (75.7% compared to 57.6 % in the implicit condition). Bosman et al. (2006) also reported an explicit learning advantage: teaching children (Dutch 9-year-olds) to read inconsistent loan words (e.g., *bureau*) phonetically helped cue their unpredictable spellings (more so than teaching children the conventional pronunciations of the loan words) and also led to successful

spelling generalizations with untrained items. Using comparable incidental and explicit tasks, Rastle, Lally, Davis, and Taylor (2021) showed that, when explicit instruction over symbol-to-sound and symbol-to-meaning mappings in two artificial systems was given, the vast majority of English-speaking adult participants generalized well (i.e., could accurately read aloud untrained words and draw spelling-to-meaning mappings within untrained words). On the other hand, only 21% of their participants reached the same high levels of performance in the discovery (i.e., incidental) condition.

de Bree et al. (2018) on the other hand, found an implicit learning advantage: They studied learning of a Dutch vowel-degeminaton rule (long vowels spelled with double letters in a final closed syllable lose one of their vowels when the noun is made plural), among Dutch-second graders (7–8-year-olds explicitly taught the spelling rule in school) relative to first graders (6–7-year-olds not yet taught the rule) using a spelling to dictation task. Children who had received explicit instruction of the rule in classroom (and were also one year older), did not demonstrate an overall better performance than the younger ones who had not; in fact, older children were shown to exploit various untaught (implicit) spelling cues in choosing between single and double *ee* spellings. This is against earlier work by Kemper et al. (2012) with 7-year-old Dutch children who compared implicit vs. explicit learning of the same rule and found no significant difference for trained items, but generalization over transfer items only among explicitly taught children.

In sum, results of the few studies that have directly evaluated learning effects under comparable implicit and explicit conditions do not converge. In addition, evidence from Dutch (a consistent orthography) may not generalize to English. Some of the studies use a naturalistic design that does not control for age differences in the implicit and explicitly taught groups of participants, and explicit instruction has been operationalized differently across studies. A comparison between strictly comparable incidental and explicit conditions for spelling patterns that are shown to be extracted from incidental reading exposure is overdue.

### **The current study**

Previous statistical learning effects that have been taken as evidence of graphotactic sensitivity in children (e.g., Samara & Caravolas, 2014; Samara et al., 2019) could have been underpinned by children's phonotactic sensitivity. The first goal of our study was to address this issue and test, using similar methods, the learning of purely graphotactic patterns, that is, with no phonological counterpart. A second goal was to investigate how different learning processes (statistical learning, explicit learning) contribute to this process. This was done by exploring children's and adults' learning of context-based patterns on single versus double letter usage in an artificial lexicon. Experiments 1 and 2 assessed whether implicit statistical learning processes alone suffice for learning in children and adults, and Experiment 3 directly compared children's incidental performance in one condition against one where the same patterns were explicitly taught. Finally, we explored associations between statistical and explicit learning ability and literacy performance by administering standardized tests of English word reading and spelling ability to all children.

No artificial learning experiments featuring pronounceable letter strings as stimuli have, to date, manipulated graphotactic sensitivity controlling for participants' phonotactic sensitivity: For example, the visual patterns studied by Samara and colleagues (2019; e.g., the letter *t* can follow *o* but not *e*) were pronounceable and so were the strings that embedded them. It is, thus, possible that learners were covertly pronouncing the stimuli and learned the underlying phonotactic patterns (e.g., /t/ can occur with /o/ but never with /ε/), as well as the underlying graphotactic patterns (e.g., the letter *t* can follow *o* but not *e*). To investigate if children and adults can learn purely orthographic (nonphonological) constraints on letter contexts incidentally, via statistical learning processes, in the current studies we assessed learning using homophone stimuli spelled with single versus double letters. These spellings (e.g., *dd*, *d*) map to the same sound (/d/), thus, learning when letters double is a purely graphotactic effect. Across all manipulations, singlet or doublet usage was

predicted by the identity of the preceding vowel such that, in Experiment 1, one medial vowel always predicted single consonants and one always predicted doublets, whereas in Experiment 2, each of the two possible vowels predicted some singlets/doublets.

We included skilled adult readers with ample print experience and beginning readers (6–7-year-olds) in line with previous work measuring real or experimenter-made pattern sensitivity (e.g., Cassar and Treiman’s (1997) youngest group; Hayes et al. (2006); Samara et al., 2019). Both were exposed, under identical task procedures, to these patterns in two brief sessions and learning generalizations were tested in two tasks. The first, legality judgment test (used in previous work), requires yes/no answers to unseen stimuli that either conformed to or violated the learned patterns, and in the second, *fill-in-the-blanks test*, participants are asked to construct conforming generalization nonwords by choosing one of the possible vowels to “fill-in” a consonantal frame (C\_C/C). Including this new and more naturalistic measure of graphotactic sensitivity would show how robust previous effects are.

Since a key question in this work is the role of explicit learning, we also used a post-experiment verbal questionnaire to obtain subjective reports of participants’ awareness of the experimenter included patterns, as well as to tap on their intuition as to what was driving their performance in both post-tests. The relationship between explicit awareness and artificial language learning performance is generally not well understood (Batterink, Reber, Neville, & Paller, 2015), and subjective reports are only imperfect measures of this construct, particularly in children. While we acknowledge this limitation, we use the questionnaire, as previous work (e.g., Treiman & Boland, 2017), to shed light on previous inconsistent findings regarding children and adults’ ability to report on untaught spelling patterns.

## Experiment 1

### Methods

#### *Participants*

To ensure our study was adequately powered to detect learning effects, we carried out a priori power analysis (preregistered at <https://osf.io/mn254> and detailed at <https://osf.io/qzww5/>; see also Supplementary Materials) based on the effects reported by Samara et al. (2019). We planned to recruit the resulting  $n$  of 44 participants but also used optional stopping in our Bayesian analyses, following Dienes (2016): Bayesian statistics are unaffected by the stopping rule provided their priors are data informed (i.e., as in our work) (Dienes, 2016; Rouder, 2014).

In this light, we first looked at the data for  $n = 25$ , checked for evidence either for  $H_0$  (evidence of above-chance learning in a given experiment) or the alternative hypothesis (evidence of no learning in a given experiment), and if the data were insensitive, we increased  $n$  by 10 participants until the null/alternative hypothesis were shown or up to 50 participants.

Using this approach, 35 typically developing Year 2 children (19 female, 16 male; mean age = 6.6 years,  $SD = 0.31$ ) took part in Experiment 1. They were all recruited using an opt-out procedure from a primary school in London, had no known language, hearing or vision impairments and no history of learning difficulties. All children were monolingual English speakers and had received the same amount of formal literacy tuition (2 years). Children were rewarded with stickers and a certificate. As in previous work (Samara et al., 2019), the mean reading and spelling performance in our sample was above average (mean reading = 123.4,  $SD = 9.77$ ; mean spelling = 121.4,  $SD = 12.8$ ), which is relatively typical in experimental studies with child participants.<sup>1</sup>

---

<sup>1</sup>WRAT-IV standardization is drawn following normative data collected from the US, where formal literacy instruction begins one year later relative to the UK. Thus, the standard scores reported here may overestimate the reading level of our participants relative to their age group in England (see Marinus et al., 2013).

Adult participants were recruited via Prolific and were tested online over two sessions. We anticipated some attrition in the second session of our study (due to our online methods), thus, our first sample consisted of 35 participants (i.e., 10 participants in excess of  $n = 25$ ). We obtained full data from twenty-nine adults (18 female, 11 male; mean age = 31.5 years,  $SD = 8.87$ ). We found substantial evidence for above-chance learning, thus, we did not recruit further participants. They reported being monolingual native speakers of English with no language, hearing, or vision impairments or any learning difficulties. They all provided informed consent and were paid for their participation.

### ***Materials***

**Graphotactic learning task.** We manipulated the joint probability of middle vowels and word-final consonants in monosyllabic words to induce a purely graphotactic constraint: Word-ending consonants always doubled in one context (i.e., following one of two possible vowels) and never doubled in another context (i.e., following the other possible vowel). Note that learning this constraint does not involve phonotactic sensitivity, in that word-final singlets and doublets (e.g., *d*, *dd*) map onto the same sound (/d/). Thus, vowel context is only predictive of word-ending letters, not sounds.

Stimuli were 64 pronounceable monosyllabic letter strings shown in Appendix A. They were created by combining one of four word-initial consonants ( $C_1$ : *d*, *g*, *m*, *r*), one of two vowels ( $V$ s: *e*, *u*) and one of four consonants (codas), either single or doubled ( $C_2$ : *f*, *ff*, *l*, *ll*, *s*, *ss*, *t*, *tt*). Six of the stimuli were real English words (*mess*, *dull*, *get*, *gull*, *gut*, *met*) but these were spread equally across types of test items, which controls for any bias they might bring to performance. The 64 stimuli were arranged into four lists, two of which conformed and two of which violated a novel purely graphotactic constraint. Nonwords from the pattern-conforming lists served as exposure and legal unseen test items, and nonwords from one of the two pattern violating lists served as illegal

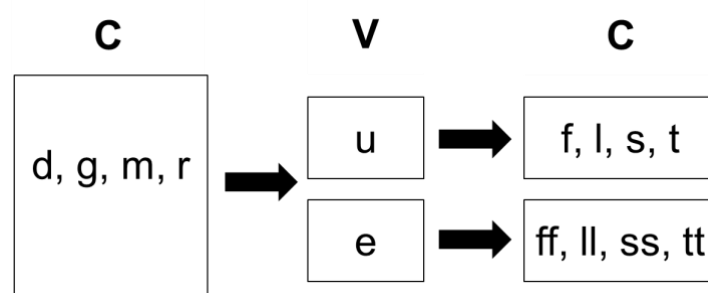
test items. Importantly, list assignment was counterbalanced between participants such that, for half of the participants word-middle *u* predicted doublets (i.e., *guff*, at test, was illegal) and for the remaining half, the same vowel predicted singlets (i.e., *guff*, at test, was legal). This counterbalancing mitigates the concern that our effects reflect children’s sensitivity to English statistical patterns.

Exposure items ( $n = 16$ ) and legal unseen items ( $n = 16$ ) conformed to the following graphotactic rule: One vowel was only followed by single consonants (e.g., in one counterbalanced list, *u* was always followed, with equal (.25) probability by *f, l, s, t*) and the other vowel was only followed by double consonants (i.e., in the same counterbalanced list, *e* was only followed with equal (.25) probability by *ff, ll, ss, tt*) (Figure 1). There were two stimuli for each legal coda (e.g., *duf, muf* and *deff, meff*). No other statistics were predictive of legality: Word-beginning consonants (C<sub>1</sub>s) co-occurred with both vowels (Vs) with .25 probability [e.g.,  $P(d, e) = P(d, u)$ ].

Illegal items ( $n = 16$ ) violated the graphotactic rule: The vowels were followed by single/doublet consonants that were not permissible (i.e., had zero probability) during exposure.

### Figure 1

*Schematic Representation of the Underlying Graphotactic Restrictions in List 1 of Experiment 1*



### *Procedure*

Child testing was carried out individually in school. The experiment ran in PsychoPy3 (Peirce et al., 2019). Adult testing was carried out online using a link distributed via Prolific

Academic. The adult experiment was designed on the Gorilla.sc platform (Anwyl-Irvine, Massonnié, Flitton, Kirkham, & Evershed, 2019) and was run on participants' own devices at home. All participants were seen in two 30-minute sessions over two consecutive days, except for 6 child participants who completed the sessions with a one-day gap.

**Graphotactic learning task.** We modelled our child-appropriate learning task based on Samara et al. (2019). At the beginning of the experiment, children were told that they were going to see written words from an alien language, called Zorib, and they would have to play games with the alien words. In session 1, the game (cover exposure task) was to detect consecutive word repetitions. In session 2, further (covert) exposure was given, followed by disclosure of the patterned nature of Zorib words. Subsequently, two new games (administered in fixed order) were (i) a task where children were asked to produce Zorib words by filling in a missing letter (“fill-in-the-blank” test), and (ii) a game where they classified new words as possible/not possible Zorib words (“legality judgment test”). Procedures for each task are detailed below.

**Exposure task.** A total of 288 Zorib “words” (144 presented in 3 blocks in each session; 9 repetitions/string in each session) were shown in the context of a one-back cover task: Participants were instructed to look at each word and press a button when repetitions occurred consecutively (16 in each session). No other instructions or feedback was given. Stimuli were presented in black in the middle of a white background and remained there until a response was given. A response was allowed only after 350ms. A fixation point (black cross, presented for 500ms) followed the response, in the middle of the screen. Word order was manipulated as follows: Consecutive stimulus repetitions occurred once for each of the 16 strings and no more than 6 times in each block. All other stimuli appeared at random and no other doubles were allowed.

***Fill-in-the-blanks task.*** A fill-in-the blanks task was devised to measure pattern sensitivity as reflected in participants' ability to choose the appropriate context (vowel) to create legal words in Zorib language. In each trial, participants saw (a) novel word frames consisting of a consonant, an underlined blank space for the missing middle vowel, and a word-final consonant (e.g., *r\_ll*) and (b) below the frame, the two vowels used during exposure (*e*, *u*). The experimenter explained that their task was to drag the vowel and fill the blank to make a word that they thought possible in Zorib language. They were encouraged to use their gut feeling and were allowed to change their mind once they saw the word in full. Stimuli ( $n = 16$ ) were presented one at a time in random order. Note that choosing correct responses made the 16 legal unseen frames used in the legality judgment test.

***Legality judgment task.*** In the legality judgment task, participants were presented with novel legal unseen ( $n = 16$ ) and illegal ( $n = 16$ ) strings in randomized order and were asked to decide if each of the words could/could not exist in Zorib language and press a corresponding button accordingly. If unsure, they were encouraged to trust their intuition or "gut feel". Each string was presented in the middle of the screen and remained until a response was given. A total of 32 items were presented in a single block.

***Awareness questionnaire.*** A brief questionnaire was administered to assess whether participants were able to verbalize the graphotactic constraints governing Zorib words. If a participant reported that they noticed patterns before they were informed regarding their presence, further questions probed what patterns they thought they noticed and how they made their choices in each of the two tests.

***Literacy measures.*** Participants' reading and spelling skills were assessed using the two relevant subtests of the WRAT-IV (Wilkinson & Robertson 2006, Green form).

## Results

Data and analyses for all experiments are available at <https://osf.io/qzww5/>. Our primary method of inference was prior-informed Bayes factors (BF) (see also Samara et al., 2019) which, unlike frequentist  $p$  values, provide evidence both for and against the null. For the majority of the analyses (critically, all comparisons against chance), priors were prespecified in our preregistered plan (<https://osf.io/mn254> for Experiment 1, <https://osf.io/kz26g> for Experiment 2, and <https://osf.io/m76ck> for Experiment 3). The few occasions where the values that informed the alternative hypothesis,  $H_1$ , were not prespecified (e.g., exploratory correlations) are stated in the relevant result sections.

In the sections below, we report BF analyses alongside frequentist statistics (logistic mixed effect models).  $p$  values are included due to their familiarity but should be interpreted with caution due to potential bias caused by optional stopping. In calculating BFs, we used the prior-informed approach outlined in Dienes (2008, 2015). For each analysis, computing the evidence for  $H_1$  over  $H_0$  requires (a) a model of the data being analysed (i.e., here,  $SEs$  and betas for the relevant coefficients from the logistic mixed effect models, in log-odds space to meet normality assumptions) and (b) a model of  $H_1$ . We modelled  $H_1$  using a half-normal distribution with an  $SD$  of  $x$ , which is appropriate for small effects closer to 0 (Dienes, 2014) and determined  $x$  (i.e., expected effect sizes under  $H_1$ ) in one of the following ways: (i) Where directly relevant independent data were available (e.g., from a methodologically similar study), we used it to infer rough estimates of the expected learning effect. For example, Samara et al. (2019), study of phono-graphotactic learning with 7-year-olds was used to estimate children's legality judgment performance across all three experiments reported here ( $\beta = 0.15$ ); (ii) In occasions where the literature could not be used to draw estimates of this sort, we constrained  $H_1$  by determining roughly maximum effects. Here, we set  $x$  to be half of the maximum, given that  $x$  is the  $SD$  of the half-normal and a maximum is

approximately 2 *SD*. An example of this approach was the *BF* analyses carried out over adults' performance in the fill-in-the-blanks task. We had previously carried out a pilot study with 20 children and 20 adults ( $\beta = 0.39$  and 1.02 for each group, respectively; see [https://rpubs.com/DSingh/Phono\\_Grapho](https://rpubs.com/DSingh/Phono_Grapho)) which validated this novel task procedure following visual as well as auditory presentation to pattern-embedding stimuli: We thus estimated that fill-in-the-blanks accuracy in our study would not exceed accuracy in the pilot study ( $1.02/2 = 0.51$ ); (iii) Finally, when neither rough estimates of the expected nor a maximum effect could be specified from independent data, we determined a plausible maximum from within the data, as detailed in the sections comparing performance between experiments. As in (ii), we set  $x$  to be half of the maximum.

We interpreted values larger than 3 as substantial evidence for  $H_1$ , values less than .33 as evidence for  $H_0$ , and values between these .33 and 3 as inconclusive or weak evidence (Jeffreys, 1961). Furthermore, since our estimates of  $H_0$  involve some subjectivity, we calculated robustness regions (*RR*) for each *BF*, notated as: *RR* [ $x_1, x_2$ ]. These show the range of estimates of  $H_1$  ( $x_1$  being the smallest and  $x_2$  the largest *SD*) for which our data would support the same qualitative conclusion. They should be interpreted bearing in mind that larger values of  $x$  bias the computation to find evidence for the null, whereas smaller values are bias in favour of  $H_1$ . They were calculated by testing values of  $x$  which are reasonable given our scale, taking increments of 0.001: Specifically, 0 log odds space to 4.595 log odds space, which corresponds to odds/odds ratio of 1.041 for a comparison between chance (or guessing performance in a group) versus almost 99% perfect accuracy.

Frequentist analyses (logistic mixed effect models with the *lme4* package in R; Bates, Mächler, Bolker, & Walker, 2015) were carried out separately for each of two binary dependent variables: accuracy in the fill-in-the-blanks task and accuracy in the legality judgment task. Fixed effects were legality (within subjects), in the analyses of legality discrimination test performance,

and experiment (between subjects) in between experiment comparisons. We worked in log-odds space to meet assumptions of normality, using estimates and *SEs*. Full random-slope structure was used (that is, by-participant slopes for all experimentally manipulated within-subject effects), as recommended by Barr, Levy, Scheepers and Tily (2013). All reported models converged with BOBYQA (Bound Optimization BY Quadratic Approximation; Powell, 2009). Conclusions regarding chance performance were established by examining the significance of the model's intercept compared to the value expected under the null hypothesis, in this case, 50% correct (or log-odds of a correct response of 0). In the legality judgment task, we also included legality (centered) as a fixed effect in the models to assess whether children were better at correctly accepting legal items, if biased to say "yes", or better at correctly rejecting illegal items, if biased to say "no". Note that the models control for these tendencies.

### *Children*

Figure 2 shows the mean proportion of children's correct responses in the fill-in-the-blanks task. We predicted that, if  $H_1$  was true, children would choose the expected vowel to create permissible generalization stimuli approximately as accurately as in Singh et al. (unpublished pilot), which was .60 correct ( $\beta = 0.39$ ). Our data provided support for this hypothesis, that is, there was evidence that children were better than chance (50%) at choosing the correct vowel,  $BF = 8.72$ ,  $RR [0.05, 1.40]$  (model intercept:  $\beta = 0.25$ ,  $SE = 0.1$ ,  $z = 2.47$ ,  $p = .01$ ). The robustness region values suggest that  $H_1$  is preferred over the null for any minimal value used to model  $H_1$  and beyond the effect size we expected.

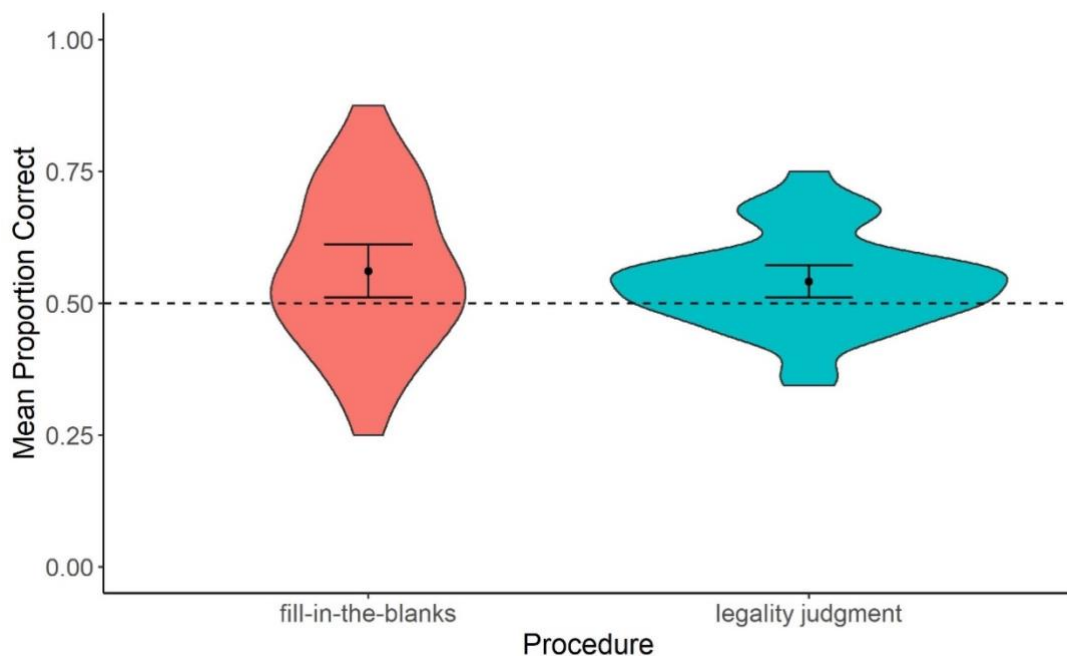
Figure 2 also shows the mean proportion of children's correct legality judgments. We predicted that, if  $H_1$  was true, children would discriminate between legal and illegal items approximately as accurately as in Samara et al. (2019) (54% correct;  $\beta = 0.15$ ). There was evidence for above (50%) chance learning,  $BF = 21.5$ ,  $RR [0.03, 2.21]$  (model intercept:  $\beta = 0.19$ ,  $SE = 0.07$ ,

$z = 2.80, p = .01$ ), that is, children were better than chance at discriminating between legal and illegal items.

In sum, children learned the graphotactic patterns. They were better-than-chance accurate both in the fill-in-the-blanks and legality judgment task.

## Figure 2

*Children's Mean Accuracy (Violin Plots With 95% Confidence Intervals) in the Fill-in-the-blanks and Legality Judgment Task in Experiment 1 (chance = .5).*



## Adults

Figure 3 shows the mean proportion of adults' correct responses in the fill-in-the-blanks and legality judgment task. For fill-in-the-blank task performance under  $H_1$ , we predicted a maximum accuracy of .72 based on our unpublished pilot data ( $\beta = 1.02/2 = 0.51$ )<sup>2</sup>. We found that adults were above (50%) chance at creating permissible generalization stimuli,  $BF = 12,455$ ,  $RR [0.05, >4.59]$ ,

<sup>2</sup>Due to an oversight, this estimate was not specified in the pre-registered plan.

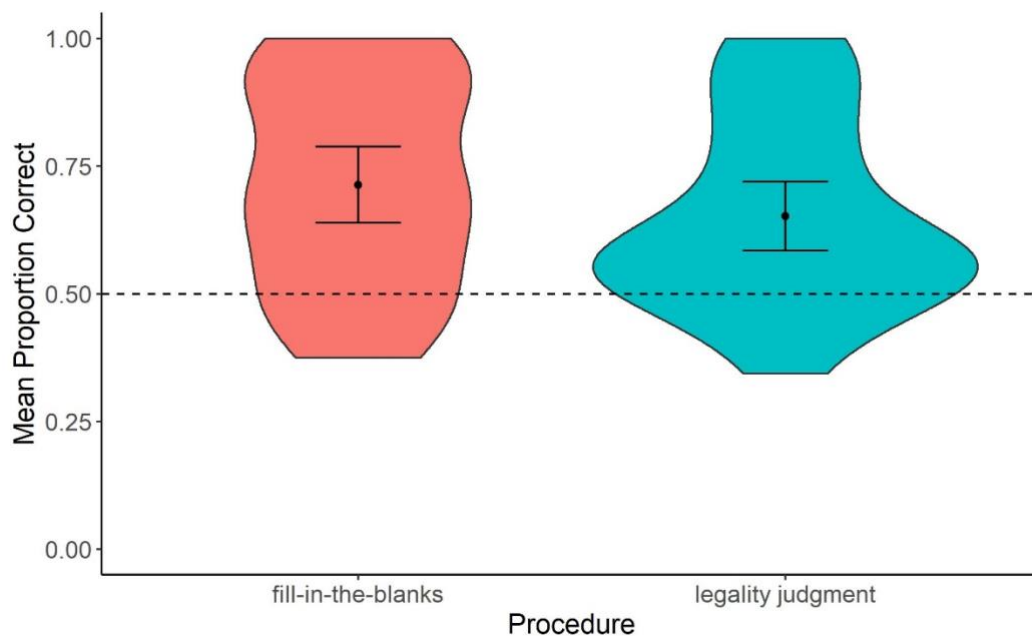
( $\beta = 1.14$ ,  $SE = 0.24$ ,  $z = 4.83$ ,  $p < .001$ ). The robustness regions suggest that  $H_1$  is preferred over the null for any minimal value used to model  $H_1$  and beyond the maximum effect we can reasonably expect to observe (99% accuracy, i.e., 4.59 in log-odds space).

For legality judgments, we predicted a maximum accuracy of .64 (model intercept:  $\beta = 0.649/2 = 0.32$ ) based on the data from the same pilot study and also confirmed that adults were above (50%) chance at discriminating between legal and illegal items,  $BF = 298.85$ ,  $RR[0.06, > 4.59]$  (model intercept:  $\beta = 0.91$ ,  $SE = 0.22$ ,  $z = 4.06$ ,  $p < .001$ ).

In sum, as with child participants, adults constructed legal items and discriminated between legal and illegal items with better than chance accuracy.

### Figure 3

*Adults' Mean Accuracy (violin plots with 95% confidence intervals) In the Fill-in-the-blanks and Legality Judgment Task in Experiment 1 (chance = .5).*



### *Awareness data*

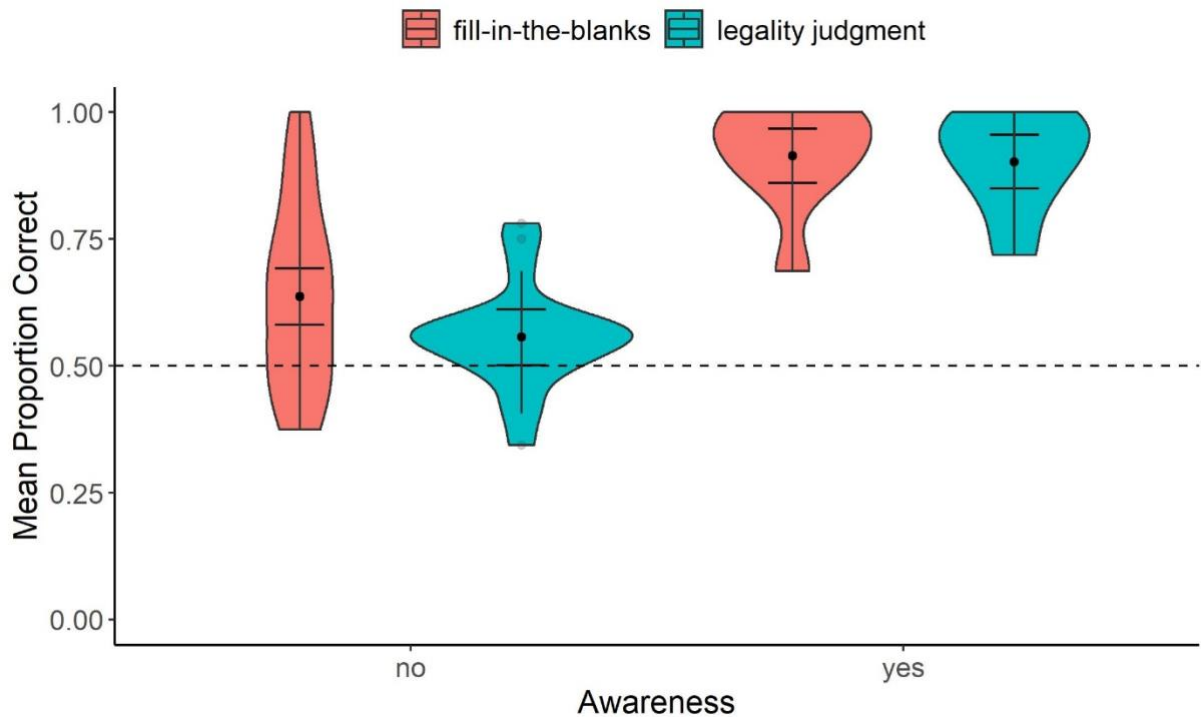
We examined participants' responses in the awareness questionnaire data to identify if they could accurately describe the novel graphotactic pattern on doubling and vowel co-occurrence. None of the children reported awareness of the patterns embedded, thus, no further analyses were carried out. Eight adult participants, on the other hand, were classified as aware on the basis of accurate descriptions such as “*Doubled end letter were preceded by the letter e. Single end letter were preceded by the letter u*”; “*Four letter words could only have e and three letter words could have u*”, “*Zorib language has an e when the last two letters are a double f*”. To investigate whether task performance was driven by this subset of aware adults, we excluded them and repeated all analyses as above.

The mean correct performance of aware and unaware adults' in the fill-in-the-blanks and legality judgment task is shown in Figure 4. From a visual inspection, it is clear that the two groups performed qualitatively differently (there is no overlap in confidence intervals between groups), and that aware participants' accuracy was close to ceiling.

For fill-in-the-blanks performance, there was evidence of above-chance learning (maximum predicted accuracy of .72;  $b = 0.51$ )  $BF = 83.64$ ,  $RR[0.07, > 4.59]$  (model intercept:  $\beta = 0.62$ ,  $SE = 0.19$ ,  $z = 3.3$ ,  $p < 0.001$ ), and so was for legality judgment performance (maximum predicted accuracy of .64;  $b = 0.32$ ),  $BF = 19.74$ ,  $RR [0.09, 1.68]$  (model intercept:  $\beta = 0.31$ ,  $SE = 0.11$ ,  $z = 2.76$ ,  $p = .01$ ).

### **Figure 4**

*Adults' Mean Accuracy (violin plots with 95% confidence intervals) by Awareness Status (aware, unaware participants) in the Fill-in-the-blanks and Legality Judgment Task in Experiment 1 (chance = .5).*



In sum, some adults demonstrated explicit awareness of the patterns embedded in the stimuli, and these participants did show markedly better performance in both tests of performance. Nevertheless, even when these participants were removed from the analysis, the remaining unaware participants were above chance as a group in both tests.

### Discussion

In Experiment 1, we investigated children and adults' ability to learn novel purely graphotactic constraints incidentally. We created stimuli ending with either single or double letters (homophones) and manipulated the context in which they occurred, such that all consonants always doubled following one of the two vowels and never following the other. We tested generalizations over this pattern by asking participants to discriminate between novel legal and illegal items (as in previous work by Samara and colleagues), and in a more naturalistic production task: Participants were presented with generalization (unseen) word frames and were asked to fill in a missing vowel,

choosing from two alternatives, in order to create permissible generalization words in Zorib language.

In both tasks, we found clear evidence of learning. Children and adults learned the nonphonological novel graphotactic constraints, which they used to judge and produce novel unseen test items, that is, performance cannot reflect the ability to memorize the 16 exposure items. None of the children reported any awareness of the patterns, but some adults ( $n = 8$ ) did (i.e., they were able to describe them when prompted at the end of the experiment). However, repeating the analysis without them confirmed that above-chance performance in the task was not driven solely by aware adults.

In sum, Experiment 1 goes beyond previous research by demonstrating purely graphotactic learning that cannot be explained by phonotactic sensitivity. One potential limitation, though, is that the patterns in this experiment were relatively simple compared to those, for example, studied by Hayes and colleagues (e.g., Hayes et al., 2006). It is therefore unclear whether our learning demonstration scales up to the challenge of learning real orthographic patterns. The somewhat simplified nature of pattern in our study might also explain why almost a third of adults could clearly articulate them. We address this limitation in a second experiment whereby we investigate children's and adults' ability to learn novel purely graphotactic constraints that are more complex, in that both vowels can be followed by singlets and doublets.

## Experiment 2

### Method

#### *Participants*

Twenty-five typically developing Year 2 children (16 female; 9 male; mean age = 6.8 years,  $SD = 0.39$ ) and 35 adults (15 female; 18 male; mean age = 31.7 years,  $SD = 9.08$ ), all monolingual

native speakers of English, took part in Experiment 2. Decisions regarding optional stopping, recruitment, consent, and compensation processes were as in Experiment 1. Exclusion criteria were also identical to Experiment 1. The mean reading standard score for children was 120 ( $SD = 9.15$ ) and the mean spelling standard score was 121 ( $SD = 15.3$ ). All participants completed two experimental sessions over two consecutive days.

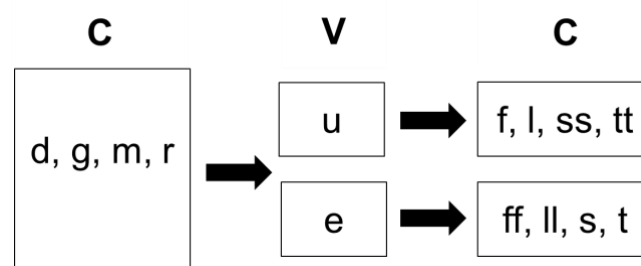
### *Stimuli and Procedure*

We manipulated the joint probability of middle vowels and word final consonants within homophonic CVC/C strings, but unlike Experiment 1, both vowels were followed by double and single letters (Figure 5). Specifically, Vowel1 was followed (i) by two of the consonants as doublets (ii) and by the other two consonants as singlets. That is, in one counterbalanced condition, *e* was always followed with equal (.50) probability by *ff*, *ll*, but not by *ss*, *tt*, and, in the same counterbalanced condition, *e* was always followed with equal (.50) probability by *s*, *t*, but not by *f*, *l*. The opposite held true for Vowel2 (Figure 5). No other statistics were predictive of legality: Word-beginning consonants ( $C_1s$ ) co-occurred with both vowels ( $Vs$ ) with .25 probability.

As in Experiment 1, the 64 pronounceable monosyllabic letter strings shown in Appendix B (28 nonwords; 4 English words: *met*, *gut*, *get*, *dull*) were arranged into three lists whose order was counterbalanced across participants.

### **Figure 5**

*Schematic representation of the underlying graphotactic restrictions in list 1 of Experiment 2*



All other aspects of the experimental task and our procedure were the same as in Experiment 1.

## Results

### *Children*

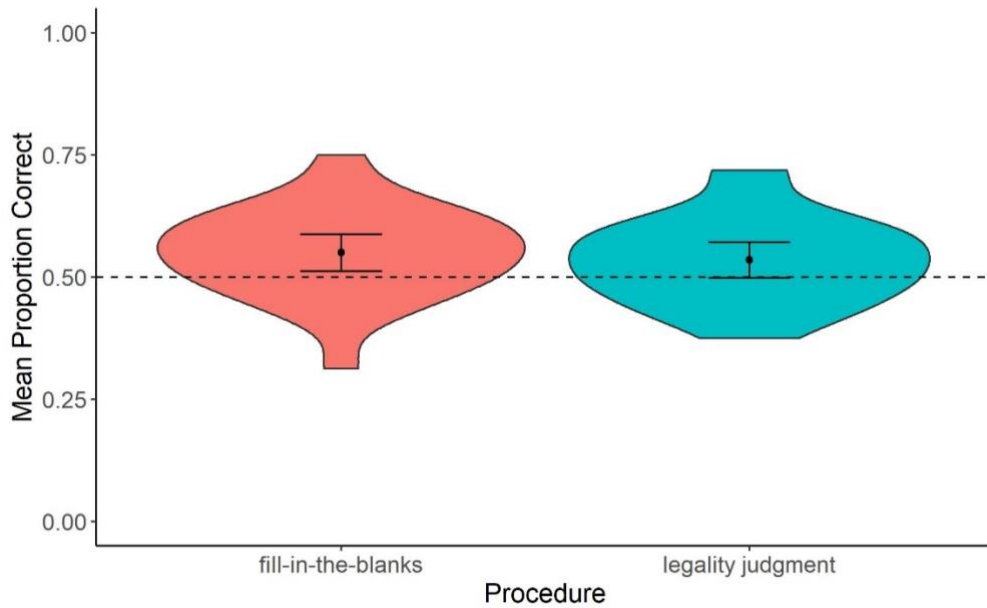
Figure 6 shows the mean proportion of children's correct responses in the fill-in-the-blanks task. We predicted that, if  $H_1$  was true, children would perform approximately as accurately as in our pilot study ( $\beta = 0.39$ ). The BFs confirmed learning was above chance (i.e., children chose the expected vowel with better than chance (50%) accuracy),  $BF = 3.13$ ,  $RR [0.08, 0.42]$  (model intercept:  $\beta = 0.2$ ,  $SE = 0.1$ ,  $z = 2$ ,  $p = .05$ ).

Figure 6 also shows children's mean correct performance in the legality judgment task. For  $H_1$ , we predicted that children would discriminate between legal and illegal items approximately as accurately as in Samara et al. (2019) ( $\beta = 0.15$ ) and the evidence supported  $H_1$  (above chance (50%) discrimination),  $BF = 5.20$ ,  $RR [0.05, 0.43]$  (model intercept:  $\beta = 0.16$ ,  $SE = 0.08$ ,  $z = 2.12$ ,  $p = .03$ ).

In sum, children learned the more complex graphotactic patterns, as evidenced by their performance in both tests.

### **Figure 6**

*Children's mean accuracy (violin plots with 95% confidence intervals) in the fill-in-the-blanks and legality judgment task in Experiment 2 (chance = .5).*



### *Adults*

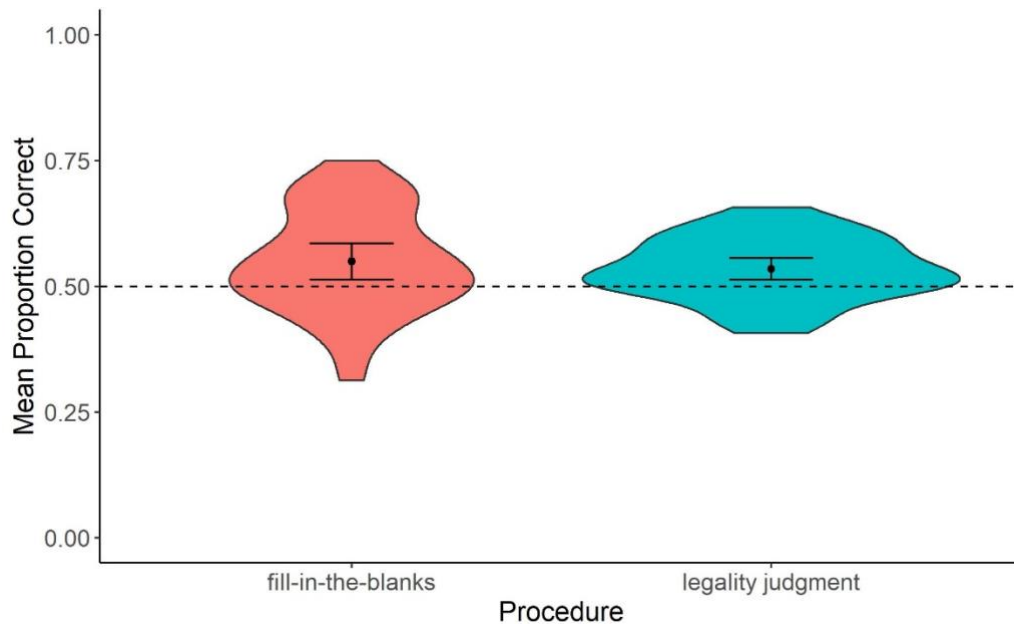
Unlike Experiment 1, none of the adults gave informative descriptions of the patterns that were introduced in the exposure phase of the experiment, thus, they were all included in the analyses reported below. Their fill-in-the-blanks and legality judgment task performance is shown in Figure 7. For the fill-in-the-blanks task, our prediction for  $H_1$  was that adults would choose the appropriate vowel in order to create permissible generalization stimuli as accurately as in our pilot study ( $\beta = 0.51$ ), and this was supported by the analyses,  $BF = 3.97$ ,  $RR [0.05, 0.70]$  (model intercept:  $\beta = 0.2$ ,  $SE = 0.09$ ,  $z = 2.26$ ,  $p = .02$ ).

Our prediction for  $H_1$  regarding performance in the legality judgment task was that adults would discriminate between legal and illegal items at most as accurately as in our pilot study (model intercept:  $\beta = 0.32$ ). Again, we found evidence for this hypothesis in terms of  $BF$  (5.18,  $RR [0.04, 0.62]$ ) (model intercept:  $\beta = 0.16$ ,  $SE = 0.07$ ,  $z = 2.31$ ,  $p = .02$ ).

In sum, the evidence from both tasks suggests that adults learned the graphotactic patterns of Experiment 2.

### **Figure 7**

*Adults' mean accuracy (violin plots with 95% confidence intervals) in the fill-in-the-blanks and legality judgment task in Experiment 2 (chance = .5).*



### ***Experiment 1 versus 2 comparison***

The results of Experiment 1 and 2 consistently suggest that children and adults pick up on novel graphotactic constraints from brief incidental exposure. Above-chance learning was seen when vowel identity predicted singlet/doublet occurrence (i.e., one vowel was consistently followed by doublets and the other was not) but also when each vowel predicted both word-ending singlets and word-ending doublets (half of each type). We hypothesized that learning the former pattern would be easier and sought to directly compare whether pattern complexity mediated the effects seen in Experiment 1 and 2. Analyses were run both for children and adults (after removing those in Experiment 1 who were able to verbalize the patterns and were possibly engaging explicit learning processes).

Note that for between-experiment comparisons, it is difficult to estimate using previous data the extent to which performance will differ between conditions. To overcome this issue, we set a

constraint on a likely maximum value from the data itself. Specifically, we predicted that the difference in these conditions would be, at most, equivalent to the scenario where Experiment 2 participants would be guessing (50% accuracy) and Experiment 1 participants would perform as accurately as seen in the data.

### **Children**

For fill-in-the-blanks task performance, we predicted that, if  $H_1$  was true, children in Experiment 1 would outperform children in Experiment 2 at most as much as estimated by the following difference score: 56% correct in the easier Experiment 1 less 50% (chance) performance in the harder Experiment 2 ( $\beta = 0.23$ ). This hypothesis could neither be accepted nor rejected, that is, the evidence for  $H_1$  was inconclusive,  $BF = 0.63$ ,  $RR [0, >4.59]$  (effect of experiment:  $\beta = 0.04$ ,  $SE = 0.13$ ,  $z = 0.33$ ,  $p = .74$ ).

For the legality judgment task, we predicted that a similar maximum difference score under  $H_1$ : 54% correct in the easier Experiment 1 — 50% (chance) performance in Experiment 2 ( $\beta = 0.17$ ). Once again, data were insensitive, i.e., we could not conclusively reject the  $H_1$ ,  $BF = 0.59$ ,  $RR [0, >4.59]$  (effect of experiment:  $\beta = 0.02$ ,  $SE = 0.10$ ,  $z = 0.22$ ,  $p = .82$ ).

In sum, we found no conclusive evidence that children were better at learning the graphotactic simpler patterns relative to more complex ones, but also no evidence for the null.

### **Unaware adults**

We conducted our analyses here excluding those participants in Experiment 1 who were coded as aware. As for children, we specified that, if  $H_1$  was true, adults in Experiment 1 would outperform adults in Experiment 2 in fill-in-the-blanks task at most by 64% correct in the easier Experiment 1 less 50% (chance) performance in Experiment 2 ( $\beta = 0.35$ ). There was substantial evidence for  $H_1$ ,  $BF = 7.23$ ,  $RR [0.10, 1.43]$ , i.e., adults were more accurate in choosing the correct

vowel in the simpler relative to the complex pattern condition (effect of experiment:  $\beta = 0.37$ ,  $SE = 0.16$ ,  $z = 2.30$ ,  $p = .02$ ). Similarly, for legality judgment performance, we predicted a maximum difference of 56% correct in the easier Experiment 1 — 50% (chance) performance in Experiment 2 ( $\beta = 0.21$ ), but here, we found inconclusive evidence for  $H_1$ ,  $BF = 1.18$ ,  $RR [0, >4.59]$  (effect of experiment:  $\beta = 0.12$ ,  $SE = 0.11$ ,  $z = 1.06$ ,  $p = .290$ ).

In sum, the evidence on adults' ability to learn the two types of patterns was mixed. They were better at selecting allowable word-medial vowels in the easier (Experiment 1) relative to the harder fill-in-the-blank condition (Experiment 2), but there was no evidence that pattern complexity affected their ability to correctly discriminate between legal and illegal items.

## Discussion

Experiment 2 introduced graphotactic constraints that were more complex and harder to articulate relative to those of Experiment 1. Specifically, word-medial vowels predicted the occurrence of both doublets and singlets but certain bigrams or trigrams (e.g., *uf*, *\*uff*, *uss*, *\*us*, in list 1) were never allowed. The key result is that both children and adults learned these constraints. This confirms that both age groups are capable of learning purely visual context-based patterns without any explicit instruction and after only a few minutes of exposure.

We also compared learning effects in Experiment 1 and Experiment 2, in order to investigate differences in generalization ability depending on the complexity of the pattern to be learned. For children, there was no substantial evidence that they learned the simple patterns better than the complex ones, but the  $BF$  was inconclusive: Thus, we cannot conclude that pattern complexity does not mediate learning performance. For adults, there was substantial evidence that easier patterns were learned better than hard ones in one of the two tasks (fill-in-the-blanks) and inconclusive evidence in the other task (legality discriminations). Note that aware participants (all in the easier experiment 1) were excluded from these analyses. This mitigates the concern that they

would drive the difference between conditions and the result suggests that the pattern tested in the current experiments is indeed harder for implicit learning mechanisms. Note, however, that removing aware participants does reduce power (we provide power awareness for inconclusive results in the General Discussion).

While no children were able to describe Zorib patterns, the subset of adults (28%) who reported them in the questionnaire of Experiment 1 reached close-to-ceiling performance (90% accurate) in both of our tasks. This finding tentatively suggests a positive link between the ability to verbally articulate (i.e., being aware of) the patterns and strong performance in the task. It does not, however, settle how and when precisely explicit awareness emerged. One possibility is that awareness was a by-product of learning, only emerging in the end or even only when participants reflect on their performance at test. Alternatively, aware participants may have begun engaging deliberate hypothesis testing processes during learning. Given that performance was so much higher for aware learners, the possibility that an explicit process may have been at work during the learning process raises a question with important implications for learning spelling patterns. Rather than relying on incidental learning alone, is learning more efficient when you are given explicit instructions about spelling patterns prior to print exposure? We investigate this in a final experiment with child participants (instead of adults; to counteract potential ceiling effects and make results more relevant to literacy development). We provide explicit instruction as to the simpler spelling patterns of Experiment 1, since these can be most straightforwardly verbalized as rules.

### **Experiment 3: Explicit Learning of Graphotactic Patterns**

#### **Method**

### *Participants*

Twenty-five typically developing Year 2 children (10 female; 15 male; mean age = 7.20 years,  $SD = 0.60$ ) completed two sessions on two consecutive days. The same recruitment, consent, and compensation processes were used as in Experiments 1a and 1b. As in our previous samples, the mean reading and spelling performance was above average (mean reading = 118.80,  $SD = 11.90$ ; mean spelling = 114.40,  $SD = 10.60$ ).

### *Stimuli*

The stimuli were identical to those in Experiment 1.

### *Procedure*

We replicated all aspects of the experimental procedure used in Experiments 1 and 2, except that, before exposure, children were explicitly told that written words in Zorib language adhered to a set of rules, which were described as follows: (In one counterbalanced list condition) “in Zorib language, the letter *e* is always followed by single letters (as in *rel*, *det*); and the letter *u* is followed by double letters (as in *rull*, *dutt*)”. Participants were then invited to perform the one-back task used in Experiments 1 and 2 over two sessions, followed by the fill-in-the blanks and legality posttests.

## **Results**

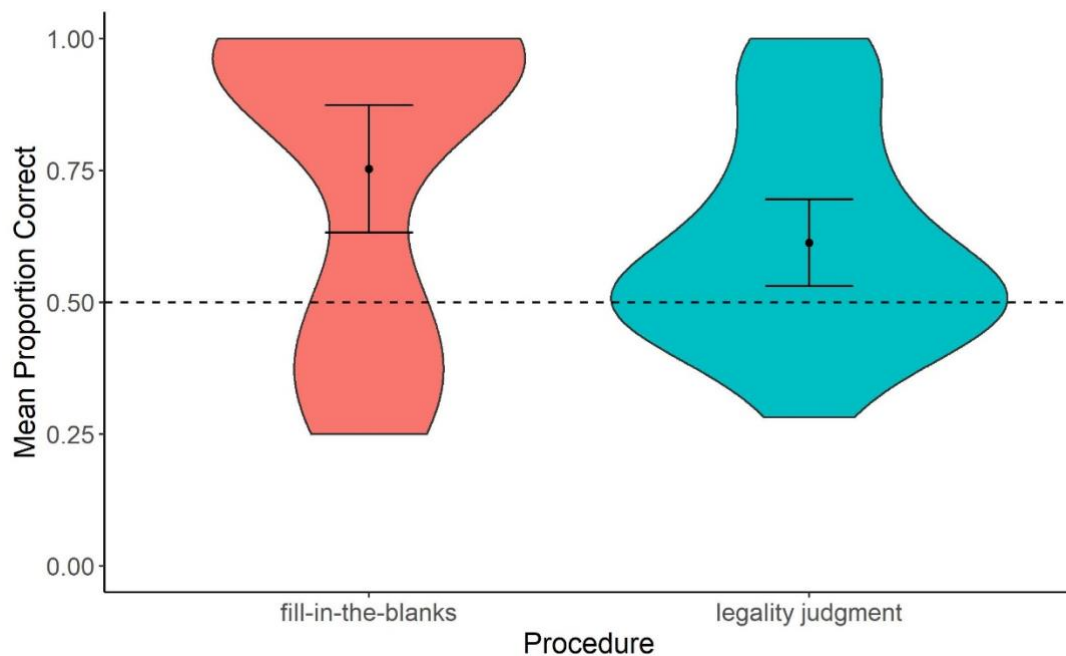
Figure 8 shows the mean proportion of children’s correct responses in the fill-in-the-blanks and legality judgment tasks. On inspection, both appear to be above chance (50%), and this was statistically confirmed.

For fill-in-the-blanks task performance, we predicted (similar to Experiment 1) that, if  $H_1$  was true, children would be as accurate as in our pilot study ( $\beta = 0.39$ ). This hypothesis was supported,  $BF = 18.25$ ,  $RR [0.17, > 4.59]$  (model intercept:  $\beta = 2.21$ ,  $SE = 0.57$ ,  $z = 3.87$ ,  $p < .001$ ).

Turning to children's legality judgments, we predicted that if  $H_1$  was true, children's accuracy will be as in Samara et al. (2019) ( $\beta = 0.15$ ). Again, we found evidence of above-chance (50%) accuracy,  $BF = 4.61$ ,  $RR [0.112, > 4.59]$  (model intercept:  $\beta = 0.74$ ,  $SE = 0.26$ ,  $z = 2.85$ ,  $p = .004$ ), despite the fact that none of the children were able to report the rules they were explicitly taught.

### Figure 8

*Children's mean accuracy (violin plots with 95% confidence intervals) in the fill-in-the-blanks and legality judgment task in Experiment 3 (chance = .5).*



### *Experiment 3 vs. Experiment 1 comparison*

In Experiment 3, we investigated 6–7-year-olds' ability to learn explicitly the novel graphotactic constraints of Experiment 1. Children were invited to play the same games as before, but were now told, at the beginning of the study, that all Zorib words were spelled according to a rule: This was explicitly taught and further illustrated via exposure to pattern-embedding instances. Results (both in the fill-in-the-blanks and legality judgment tasks) supported  $H_1$ , thus, performance

was reliably better than chance, and in fact, numerically higher relative to the condition where patterns were learned from mere exposure to the pattern-embedding stimuli. To directly investigate whether there is a relative advantage for explicit instruction above incidental learning, we ran comparisons between Experiment 1 and Experiment 3, given that the stimuli in both experiments are identical.

As in previous comparisons between experiments, we constrained  $H_1$  by means of maximum value taken from the data itself. We predicted that, if  $H_1$  was true, in fill-in-the-blanks task, children would be better at creating permissible generalization items when explicitly instructed about the patterns compared to when they learned them implicitly, at most as much as they would if they constructed legal items only when they were explicitly instructed but not when they learned implicitly ( $\beta = 0.83$ ). We found substantial evidence for  $H_1$ ,  $BF = 389$   $RR [0.104, >4.59]$  (effect of experiment:  $\beta = 1.33$ ,  $SE = 0.35$ ,  $z = 3.83$ ,  $p < .001$ ).

Similarly, for the legality judgment task, we predicted that, if  $H_1$  was true, children would be better at discriminating between legal and illegal items when explicitly instructed about the patterns compared to when they learned them implicitly, at most as much as they would if they discriminated between legal and illegal items only when they were explicitly instructed but not when they learned implicitly ( $\beta = 0.38$ ). We found substantial evidence for  $H_1$ ,  $BF = 5.93$ ,  $RR [0.13, 1.32]$  (effect of experiment:  $\beta = 0.43$ ,  $SE = 0.19$ ,  $z = 2.19$ ,  $p = .03$ ).

In sum, children's performance was stronger under explicit instructions relative to comparable incidental exposure. This was true both for performance in the fill-in-the blanks and legality judgment tasks.

### **Associations between learning and literacy performance**

There is ongoing debate regarding the relationship between learning via statistical learning processes and written (as well as spoken) language ability, due to mixed significant and

nonsignificant patterns of association shown in previous work. In our studies, a secondary goal was to explore associations between learning performance and accuracy on standardized (WRAT-4) reading and spelling performance. These analyses were exploratory and therefore not pre-registered. We addressed this question among children and, in contrast to previous work, used BF analyses to quantify evidence for both  $H_1$  (positive associations such that those who performed better in the learning tasks were also better readers/spellers) and the null (no relationship between statistical learning and literacy skills). To anticipate our results, the explicit learning task (Experiment 3) was the only condition where associations were significant, and this coefficient was used to inform the predicted relationship in Experiments 1 and 2. That is, we predicted that, if  $H_1$  was true in Experiments 1 and 2, we would find a correlation approximately as strong as that found in Experiment 3 (Fisher's  $z$ -transformed coefficient = .52). For the latter experiment, the estimate of the expected coefficient under the  $H_1$  was a small-to-medium effect ( $r = .40$ , i.e.,  $z = .42$ ), in line with coefficients reported in previous studies reporting significant associations.

Results from our correlation analyses are presented in Table 1. To sum up the results, in the incidental Experiments 1 and 2, the evidence for the association between learning performance and literacy was inconclusive in all but three occasions where we demonstrated the null (no relationship between task performance and literacy): (a) Fill-in-the-blanks performance in experiment 1 and spelling ability; and (b) fill-in-the-blanks performance in experiment 2 and spelling ability; and (c) legality performance in Experiment 2 and spelling ability. In order to maximise the available evidence, we also pooled across the two experiment datasets and explored correlations: We found substantial evidence for the null in all occasions except for the association between legality judgment performance and reading performance, where the evidence remained inconclusive. Turning to Experiment 3 (explicit learning), we found substantial evidence for a positive association (better learning scores in better readers) between legality performance and reading and spelling

ability and between fill-in-the-blanks performance and reading ability, but inconclusive evidence regarding fill-in-the-blanks performance and spelling ability.

In sum, when conclusive patterns of correlations emerged, for incidental learning, these were evidence of no relationship between learning performance and literacy; for explicit learning, they were evidence for a positive association between explicit learning and literacy ( $H_1$ ).

**Table 1**

*Correlations between accuracy in Experiment 1, 2 & 3 and WRAT reading and Spelling raw scores  
(by procedure)*

Experiment & Procedure	Statistics	Reading	Spelling
<b>Experiment 1</b>			
Fill-in-the-blanks	<i>BF</i> [RR]	0.17 [0, >4.59]	<b>0.33<sup>a</sup></b> [0, >4.59]
	<i>P</i>	.29	.95
	$z_r$ ( $SE_{z_r}$ )	-0.19(0.18)	0.01(0.18)
Legality Judgment	<i>BF</i> [RR]	0.76 [0, >4.59]	0.39 [0, >4.59]
	<i>P</i>	.35	.80
	$z_r$ ( $SE_{z_r}$ )	0.16(0.18)	0.05(0.18)
<b>Experiment 2</b>			
Fill-in-the-blanks	<i>BF</i> [RR]	0.58 [0, >4.59]	<b>0.27<sup>a</sup></b> [0, >4.59]
	<i>P</i>	.61	.61
	$z_r$ ( $SE_{z_r}$ )	0.11(0.21)	-0.11 (0.21)
Legality Judgment	<i>BF</i> [RR]	0.54 (0, >4.59]	<b>0.21<sup>a</sup></b> (0, >4.59]
	<i>P</i>	.66	.37
	$z_r$ ( $SE_{z_r}$ )	0.09(0.21)	-0.19(0.21)
<b>Experiment 1 and 2, Pooled</b>			
Fill-in-the-blanks	<i>BF</i> [RR]	<b>0.16<sup>a</sup></b> [0, >4.59]	<b>0.21<sup>a</sup></b> [0, >4.59]
	<i>P</i>	.55	.82
	$z_r$ ( $SE_{z_r}$ )	-0.08(0.13)	-0.03(0.13)
Legality Judgment	<i>BF</i> [RR]	0.65 [0, >4.59]	<b>0.18<sup>a</sup></b> [0, >4.59]
	<i>P</i>	.32	.65
	$z_r$ ( $SE_{z_r}$ )	0.13(0.13)	-0.06(0.13)
<b>Experiment 3</b>			
Fill-in-the-blanks	<i>BF</i> [RR]	<b>3.46<sup>b</sup></b> [0.20, 0.62]	1.40 [0, >4.59]
	<i>P</i>	.06	.22
	$z_r$ ( $SE_{z_r}$ )	0.40(0.21)	0.26(0.21)
Legality Judgment	<i>BF</i> [RR]	<b>9.42<sup>b</sup></b> [0.12, 2.67]	<b>5.74<sup>b</sup></b> [0.14, 1.38]
	<i>P</i>	.02	.03
	$z_r$ ( $SE_{z_r}$ )	0.52(0.21)	0.46(0.21)

<sup>a</sup> substantial evidence for  $H_0$

<sup>b</sup> substantial evidence for  $H_1$

### General Discussion

In three comparable learning experiments, we assessed children and adults' ability to pick up on purely visual graphotactic constraints between word-final single versus double consonants and letter-vowel context. Across experiments and age groups, learning was induced by exposure to pattern-embedding stimuli and generalization was assessed in two post-tests (legality judgment and fill-in-the gap task). There were two key manipulations across the three experiments presented here: First, learning was induced either incidentally (no instruction to learn the patterns; Experiment 1 & 2) or explicitly (explicit rule teaching followed by the same cover task; Experiment 3); and second, patterns varied in complexity, such that, in one experiment, vowel identity predicted word-final singlet/doublet usage (Experiment 1 and 3: "Easy" graphotactic pattern), whereas in Experiment 2 the restriction concerned specific word-middle vowel + word-ending singlet/doublet combinations, rather than doubling per se. Further to the experimental learning task, child participants were administered measures of literacy (reading and spelling task), in order to explore associations between incidental/explicit learning and literacy performance.

There were three key results. First, both children and adults incidentally learned the novel graphotactic constraints that had no phonological counterpart and consistently generalized over the test stimuli: They produced permissible spellings by selecting allowable word-medial vowels and discriminated between permissible and impermissible strings with better-than-chance accuracy. Moreover, they were able to do this for more complex rules. Second, adults who picked up on the task explicitly did better; no children could report the untaught rules but a final experiment showed that explicitly teaching children a graphotactic rule yielded significant learning effects that were stronger relative to those induced incidentally. Finally, there was only evidence of significant associations between test generalization performance and performance on reading and spelling measures under explicit conditions. Incidental learning performance was either unrelated to

participants' reading/spelling ability or data were inconclusive. We discuss each of these findings in turn below.

### **Implicit learning of visual graphotactic spelling patterns**

We provide the first strong demonstration that visual statistical learning processes underlie the learning and generalization of purely graphotactic spelling patterns in children, controlling for the possibility that statistical learning accounts for children's sensitivity to spoken language patterns. Previous work measuring children's sensitivity to spelling patterns within a familiar language has shown that children prefer stimuli that conform to such regularities. For example, Hayes et al. (2006) showed that English-speaking children prefer nonwords embedding doublets after single-vowel than double-vowel spellings (e.g., *saff* > *\*saf*, following *Jeff/bedding/pull* versus *deaf/heading/sail*) and have explained these effects in terms of statistical sensitivity to orthographic properties. However, the learning mechanisms at play were postulated, not directly shown; it is also hard—as with all studies in this line of work—to establish whether developmental patterns of improvement were due to maturational differences in learning ability or differences in the amount of children's exposure to print.

We address these limitations by manipulating patterns in a well-controlled learning experiment, maintaining the use of pronounceable stimuli to draw close parallels to how people learn existing writing systems. The same approach was taken by Samara and colleagues (Samara & Caravolas, 2014; Samara et al., 2019) who manipulated patterns of letter co-occurrence within consonant-vowel-consonant strings and interpreted their learning effects as evidence for graphotactic sensitivity. However, their stimuli and patterns were all pronounceable, thus, children's apparent graphotactic sensitivity could be underpinned by phonotactic learning (Chambers et al., 2003; Onishi et al., 2002).

It is also possible to use fully artificial languages to this effect, however, there are important methodological considerations at play. Firstly, the artificial symbols need to be appropriate for use with children but in order to serve as evidence for purely graphotactic learning, they cannot map directly onto verbal labels (e.g., rectangles, triangles, or other easily named shapes). Secondly, children are likely to need multiple training sessions to demonstrate learning over novel sequences comprising unfamiliar symbols, as seen in artificial spoken language experiments with novel items (e.g., see Wonnacott, 2011; Samara, Smith, Brown, & Wonnacott, 2017). As a potential solution, future experiments may explore the possibility of embedding patterns within stimuli comprising familiar letters and novel symbols. We are currently investigating this possibility in pilot work.

To date, only a handful of studies have used fully artificial systems (Chetail, 2017; Lelonkiewicz, Ktori, & Crepaldi, 2020; Nigro et al., 2006; Vidal, Viviani, Zoccolan, & Crepaldi, 2021) but only two have used pseudoletter strings or child participants to test the learning of graphotactic patterns. Chetail (2017) taught adults an unfamiliar language (words comprising characters from the Phoenician Moabite alphabet that have no phonological counterpart for French-speaking participants) and showed learning of positional and letter co-occurrence frequencies in generalization stimuli. The study did not, however, include child participants and thus cannot provide support against the claim that children are insensitive to these patterns early in spelling development (Frith, 1985; Gentry, 1982). Nigro et al. (2016) tested 8-year-olds, but the novel patterns were unconditional, i.e., restricted letter positions rather than letter combinations. This ability mirrors naturalistic evidence that children's invented spellings often conform to positional constraints (e.g., "letters begin with *ck*") (Treiman, 1993) but raises the question of whether children are also able to learn context-based patterns which are harder to learn from familiar and novel languages (Samara & Caravolas, 2014; Hayes et al., 2006).

In our work, we addressed this issue by using homophone letters (e.g., *dd* and *d* map onto the same sound). One concern with using the children's native alphabet is that children may bring

knowledge from English patterns which interfere with the novel graphotactic learning task. For example, some of the stimuli violate an existing English pattern—*f* would be doubled in these stimuli so those stimuli with *ff* may be easier to learn than those with *f*, where an existing pattern has to be unlearned. Critically, however, as explained under Procedure, such items are counterbalanced as legal and illegal between participants, so that such relationship with English stimuli may add noise but cannot underpin our effects. One further potential concern was that, even though homophone letters are pronounced the same by skilled readers, children might have covertly pronounced them differently, implicating phonotactic learning in their performance. We therefore conducted a brief online study<sup>3</sup> and clearly demonstrate that this was not the case: 14 children (6–7 years old) read aloud the full set of nonwords (presented one at a time) that comprised the stimuli in Experiment 2. Subsequently, two native English speakers who were blind to the purpose of the experiment, listened to all of the recordings that were grouped (unbeknown to them) into ‘homophone’ pairs (e.g., a child’s responses for *det* and *dett*) versus foil pairs (within-participant recordings were mixed at random). Of the 448 homophone trials (32 per participant), 85% of them were identified by both coders as the ‘same’ word, 11% were poor quality recordings (e.g. one of the recordings cut off prematurely or was missing), and only 4% were cases where the child actually produced two different words. Importantly, these were all idiosyncratic pronunciations made by individual children and mainly reflected differences in onset pronunciations (e.g., *ges* vs. *yes*) rather than in different pronunciations for the codas, mitigating the concern that homophone letters may have been pronounced differently by the children in the main experiments. One final point we acknowledge is that even if the difference between the conditioned graphotactic patterns has no phonological counterpart, the conditioning graphotactic environment—that is, the vowel—is itself pronounceable. As noted above, future experiments could address this by using an artificial

---

<sup>3</sup> data can be found at <https://osf.io/h9gdr/>

lexicon of fully unpronounceable stimuli. In terms of the current study, our contribution is to establish that children can learn constraints on the occurrence of different graphotactic patterns even when the graphotactic differences are not supported by correlated phonological differences, and thus learning of the constraints cannot depend on learning phonotactics. In sum, we have thus demonstrated above-chance learning of visual graphotactic constraints which cannot be underpinned by phonotactic learning<sup>4</sup>. This suggests that literacy—a recent human accomplishment from an evolutionary perspective—is underpinned by statistical learning mechanisms. These have been implicated in generalizations over both sequentially presented patterns (within spoken stimuli, Saffran et al. 1996; within visual stimuli; Fiser & Aslin, 2002) as well as simultaneously presented visual (spatial) patterns (Fiser & Aslin, 2001), and are thought to play a key role in child first language acquisition (Wonnacott, 2013). Their role in literacy acquisition has been previously speculated (e.g., Kessler, Pollo, Treiman, & Cardoso-Martins, 2013; Treiman et al., 2017). We provide the first direct evidence for this in relative beginner spellers tested on context-based graphotactic patterns. The demonstration of graphotactic sensitivity in 6–7-year-olds is in contrast to literacy models which predict that young children are not sensitive to letter patterns, morphological information, and other advanced sources of knowledge until they enter the final (correct or fully alphabetic) stage of spelling development (Frith, 1985; Gentry, 1982).

Further to this important demonstration of graphotactic sensitivity across two patterns, we investigated the effect of complexity on generalization ability in childhood and adulthood. We argued that learning to predict word-final singlet/doublet usage from previous vowel context was easier than learning specific combinations between word middle vowels and subsequent single/double letters. This is because, while the underlying joint probabilities of vowels, word-ending letters were the same across conditions (1 in 4 for each vowel and final (single/double) letter

---

<sup>4</sup> We thank an anonymous reviewer for all the points we address in this paragraph.

combination), there is an additional (higher-level) joint probability statistic underlying the stimuli used in Experiment 1. Namely, participants could pick up on the joint probability of 1 between Vowel1 and singlets, as opposed to the joint probability of 0 between the same vowel and doublets. In Experiment 2, the joint probability of Vowel1 and singlets was .50 and so was the joint probability of Vowel1 and doublets.

For adults, we found substantial evidence that easier patterns were learned better than hard ones in one of the tasks, namely, the fill-in-the-blanks task. Note that our analyses included only unaware participants, indicating that this difference is not due to the rule in Experiment 1 being easier to articulate and, thus, pick up explicitly: Notably, this is easily verbalized as “*e* is always followed by singlets, never doublets”. The evidence regarding adult performance in legality judgments and children’s performance in both tasks was on the other hand inconclusive. Thus, we cannot interpret null findings for children as demonstrating that they learn simple and complex patterns similarly. Supplementary analyses (assuming that the error term would reduce in proportion to  $\sqrt{SE}$ ; see <https://osf.io/qzww5/>) suggest that to establish  $H_1$  (i.e., demonstrate better learning for the easier patterns) 4000 more child and 130 more adult participants would be needed). In sum, our work tentatively suggests that the easier patterns were better learned by implicit learning mechanisms than the harder patterns. Ideally, this result should be replicated in a larger sample. What the current results clearly establish is that the complex patterns *can* be learned, as well as the simple ones, even under incidental, implicit learning conditions by children and adults.

One further contribution of our work is the methodological development and validation of a production fill-in-the-blanks task as a new measure of artificial pattern sensitivity. Unlike the legality judgment task (which we also employed for consistency with previous work) production performance simulates more closely what children and adults do in naturalistic situations. Thus, our study goes beyond previous work by showing that children’s knowledge of novel orthographic constraints generalizes, to some extent, to their own (partial) written productions. Other possible

avenues for methodological contributions to artificial orthographic learning research is the development and validation of online tasks (e.g., reaction-time based tasks and neural measures such as event-related potentials) that are more indirect (and possibly more sensitive) measures of tacit knowledge (Batterink, Cheng, & Paller, 2016; Siegelman, Bogaerts, Kronenfeld, & Frost, 2018).

### **Explicit learning of visual graphotactic spelling patterns**

In Experiment 3, we assessed children's ability to learn novel graphotactic constraints under explicit task instructions, and contrasted—for the first time, to our knowledge—the relative effectiveness of explicit and incidental processes involved in orthographic knowledge acquisition using artificial language methods. Much of previous work that has sought to address questions related to the effectiveness of implicit and explicit orthographic pattern learning suffers from methodological weaknesses. For example, only two orthographic learning studies (Kemper et al., 2012; Sobaco et al., 2015) have controlled for idiosyncratic task differences by using implicit and explicit versions of the same tasks.

In our well-controlled experiments, learning patterns explicitly was clearly advantageous for performance in both of our tests. This demonstration converges with the practice of teaching spelling patterns, even if they are partially predictive, from early on in literacy instruction. Note that our findings primarily concern typical populations but are also relevant for poor spellers, including children diagnosed with dyslexia. Future experiments should investigate whether the advantage for explicit over incidental graphotactic learning also holds in these populations, as suggested by previous educational research (Wanzek et al., 2006). Graphotactics are an interesting case as they have been shown to affect nonword spellings produced by adults ranging from poor to good spelling ability (Treiman & Boland, 2017). This is in contrast to work documenting statistical learning deficits in dyslexic children (Pavlidou, Kelly, & Williams, 2010; Ise, Arnoldi, Bartling, &

Schulte-Körn, 2012; Vicari et al., 2005; see, however, Nigro et al., 2016). We believe that our work combining insights from the literatures on real graphotactic sensitivity and the statistical learning of artificial languages has great potential to shed light on such conflicting findings.

It would be also interesting to explore whether the advantage of explicit instruction over incidental learning holds for patterns that can be less clearly explained/verbalized, considering the abundance of English useful vocabulary statistics that are too complex to verbalize, both in childhood and adulthood (Kessler, 2009). This would provide a direct test of Reber, Walkenfeld, and Hernstadt's (1991) claim that searching for complex grammatical rules and linguistic patterns impedes learning. In our work, we chose rules that are easier to verbalize as a starting point, but it may be possible to teach more complex and harder to describe rules, as those used in Experiment 2, in parts and examining whether learning advantages occur. Finally, future work needs to explore more systematically what specifically makes explicit instruction so beneficial in experiments (and real-life conditions). Does it suffice to bring the pattern embedding nature of the stimuli to participants' attention or does the rule need to be taught? Would the benefits have been eliminated if the rule was not exemplified by further exposure? These are questions with important implications for classroom instruction.

### **Associations between learning performance and literacy skill**

One final exploratory question addressed in our work with developing learners (6–7-year-olds) concerned the widely hypothesized link between statistical learning processes and literacy performance. A few previous studies have empirically demonstrated that variations in lab-based statistical or implicit learning skill are related to literacy performance among typically developing children and adults (most notably, Arciuli & Simpson, 2012; Frost et al., 2013; Spencer et al., 2015), but others have not (Schmalz et al., 2019; West et al., 2018).

In our study, we did not replicate the evidence for positive association under incidental learning conditions. While some of the relevant correlations in each implicit experiment were inconclusive when data was pooled across the two experiments, we conclusively showed that variations in learning skill were unrelated to variations in spelling ability. Does this null finding suggest, more generally, that statistical learning skill is unrelated to literacy performance? This is one possibility, pending future replication and extension of our work using other measures of literacy and literacy-related (e.g., phonological awareness skill) skills. Another possibility is that null findings reflect our methods, in that real graphotactic knowledge may have competed with graphotactic manipulations in our experiments. In this case, robust internalized representations of English graphotactics among better learners could cancel out learning of the experimental constraints.

Limitations in our work include that we used single measures of each literacy construct, as well as the fact that we did not seek to establish strong psychometric properties (validity and reliability) for our task. This is because exploring correlations was a secondary (exploratory) aim in our work, thus, we cannot exclude the possibility that our null (as well as the inconclusive) result may reflect the psychometric issues pointed out by Siegelman et al. (2017) and West et al. (2018). A strength of our work, on the other hand, is the inclusion of correlational Bayes factors that we believe should complement all future work reporting nonsignificant associations. In our case, they clearly demonstrate that nonsignificant correlations do not necessarily constitute evidence for the null, as often interpreted in the literature. For example, a recent study by Qi, Sanchez Araujo, Georgan, Gabrieli, and Arciuli (2019) reported significant associations between auditory but not visual statistical learning ability and reading fluency and took this as evidence that “hearing is more important than seeing” for literacy development. In line with Dienes (2014), we caution against these statements on the basis of frequentist results alone.

Turning to the pattern of correlations seen with measures of explicit learning, we obtained substantial evidence for  $H_1$  (positive link between statistical learning and literacy performance) for measures and variations in reading/spelling in all but one occasion that involved the fill-in-the-blank task (possibly due to reduced item-based power in this task; 36 more participants needed to demonstrate  $H_1$ , assuming that the error term would reduce in proportion to  $\sqrt{SE}$ ; see <https://osf.io/qzww5/>). These positive associations strengthen the view that explicit learning contributes to learning to read and spell. We acknowledge that our data are correlational and not evidence for causation, hence, it is, for example, possible that children with better reading skills *also* have stronger language skills that help them better understand the task instructions. Last, while we maintain that our analyses are exploratory, the contrast in the implicit and explicit correlational results is intriguingly consistent with a general theory of implicit learning being less susceptible to individual differences than explicit learning due to its evolutionary precedence (Reber, 1989; Reber et al., 1991). This has been most strongly shown in terms of associations with memory and intellectual ability (Kaufman et al., 2010; Gebauer & Macintosh, 2007).

### Conclusion

In conclusion, we have demonstrated that visual statistical learning processes support the learning of graphotactic patterns that do not correlate with phonology. Under explicit rule instructions, learning improves. An implication of these findings for spelling instruction is that early incidental exposure to print should be encouraged, although explicit teaching, at least of the patterns studied here is altogether more effective. An interesting challenge for future lab-based work would be to move away from describing learning as purely implicit or explicit. It is more likely that these mechanisms interact in complex ways over the course of becoming literate (Steffler, 2001; Karmiloff-Smith, 1992). It would be also interesting to explore whether learners register associations explicitly and integrate them implicitly or vice versa (Yang & Li, 2012).

### Acknowledgements

We would like to thank the schools and children who took part in these studies. We would also like to thank Vanessa Fortune and Briana Lelliott for coding the online reading aloud data we mention in the General Discussion.

### References

- Anwyl-Irvine, A. L., Massonnié, J., Flitton, A., Kirkham, N., & Evershed, J. K. (2019). Gorilla in our midst: An online behavioral experiment builder. *Behavior Research Methods*, *52*(1), 388–407. <https://doi.org/10.3758/s13428-019-01237-x>
- Arciuli, J., & Simpson, I. C. (2012). Statistical learning is related to reading ability in children and adults. *Cognitive Science*, *36*(2), 286–304. <https://doi.org/10.1111/j.1551-6709.2011.01200.x>
- Arciuli, J., von Koss Torkildsen, J., Stevens, D. J., & Simpson, I. C. (2014). Statistical learning under incidental versus intentional conditions. *Frontiers in Psychology*, *5*(JUL), 1–8. <https://doi.org/10.3389/fpsyg.2014.00747>
- Barr, D. J., Levy, R., Scheepers, C., & Tily, H. J. (2013). Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of Memory and Language*, *68*(3). <https://doi.org/10.1016/j.jml.2012.11.001.Random>
- Bates, D., Mächler, M., Bolker, B. J., & Walker, S. C. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, *67*, 1–48. <http://dx.doi.org/10.18637/jss.v067.i01>
- Batterink, L. J., Cheng, L. Y., & Paller, K. A. (2016). Neural measures reveal implicit learning during language processing. *Journal of Cognitive Neuroscience*, *28*(10), 1636–1649. <https://doi.org/10.1162/jocn.a.00985>

- Batterink, L. J., Reber, P. J., Neville, H. J., & Paller, K. A. (2015). Implicit and explicit contributions to statistical learning. *Journal of Memory and Language*, 83, 62–78.  
<https://doi.org/10.1016/j.jml.2015.04.004>
- Bosman, A. M. T., van Hell, J. G., & Verhoeven, L. (2006). Learning the spelling of strange words in Dutch benefits from regularized reading. *Journal of Educational Psychology*, 98(4), 879–890. <https://doi.org/10.1037/0022-0663.98.4.879>
- Butyniec-Thomas, J., & Woloshyn, V. E. (1997). The effects of explicit-strategy and whole-language instruction on students' spelling ability. *The Journal of Experimental Education*, 65(4), 293–302. <https://doi.org/10.1080/00220973.1997.10806605>
- Camerer, C. F., Dreber, A., Holzmeister, F. *et al.* (2018). Evaluating the replicability of social science experiments in *Nature and Science* between 2010 and 2015. *Nature Human Behaviour*, 2, 637–644. . 1–26. <https://doi.org/10.1038/s41562-018-0399-z>
- Caravolas, M., Kessler, B., Hulme, C., & Snowling, M. (2005). Effects of orthographic consistency, frequency, and letter knowledge on children's vowel spelling development. *Journal of Experimental Child Psychology*, 92(4), 307–321. <https://doi.org/10.1016/j.jecp.2005.08.001>
- Caravolas, M., Lervåg, A., Mousikou, P., Efrim, C., Litavský, M., Onochie-Quintanilla, E., ... & Seidlová-Málková, G. (2012). Common patterns of prediction of literacy development in different alphabetic orthographies. *Psychological Science*, 23(6), 678–686.  
<https://doi.org/10.1177/0956797611434536>
- Carrillo, S., & Alegría, J. (2014). The development of children's sensitivity to bigram frequencies when spelling in Spanish, a transparent writing system. *Reading and Writing*, 27, 571–590.  
<https://doi.org/10.1007/s11145-013-9459-y>

- Cassar, M., & Treiman, R. (1997). The beginnings of orthographic knowledge: Children's knowledge of double letters in words. *Journal of Educational Psychology*.  
<https://doi.org/10.1037/0022-0663.89.4.631>
- Chambers, K. E., Onishi, K. H., & Fisher, C. (2003). Infants learn phonotactic regularities from brief auditory experience. *Cognition*, 87, B69–B77. <https://doi.org/10.1016/S0>
- Chetail, F. (2017). What do we do with what we learn? Statistical learning of orthographic regularities impacts written word processing. *Cognition*, 163, 103–120.  
<https://doi.org/10.1016/j.cognition.2017.02.015>
- de Bree, E., Geelhoed, J., & van den Boer, M. (2018). Overruled!: Implicit cues rather than an orthographic rule determine Dutch children's vowel spelling. *Learning and Instruction*, 56, 30–41. <https://doi.org/10.1016/j.learninstruc.2018.03.006>
- Dienes, Z. (2008). Understanding psychology as a science: An introduction to scientific and statistical inference. *Palgrave Macmillan*.
- Dienes, Z. (2014). Using Bayes to get the most out of non-significant results. *Frontiers in Psychology*, 5(July), 1–17. <https://doi.org/10.3389/fpsyg.2014.00781>
- Dienes, Z. (2015). How Bayesian statistics are needed to determine whether mental states are unconscious. In M. Overgaard (Ed.), *Behavioural Methods in Consciousness Research* (pp. 199–220). Oxford: Oxford University Press.
- Dienes, Z. (2016). How Bayes factors change scientific practice. *Journal of Mathematical Psychology*, 72, 78–89. <https://doi.org/10.1016/j.jmp.2015.10.003>
- Fiser, J., & Aslin, R. N. (2001). Unsupervised statistical learning of higher-order spatial structures from visual scenes. *Psychological Science*, 12, 499–504. <https://doi.org/10.1111/1467-9280.00392>

- Fiser, J., & Aslin, R. N. (2002). Statistical learning of higher-order temporal structure from visual shape sequences. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 28, 458–467. <https://doi.org/10.1037/0278-7393.28.3.458>
- Frensch, P. A., & Rüniger, D. (2003). Implicit learning. *Current Directions in Psychological Science*, 12(1), 13–18. <https://doi.org/10.1111/1467-8721.01213>
- Frith, U. (1985). Beneath the surface of developmental dyslexia. In *Journal of Chemical Information and Modelling* (Vol. 53). <https://doi.org/10.1017/CBO9781107415324.004>
- Frost, R., Siegelman, N., Narkiss, A., & Afek, L. (2013). What predicts successful literacy acquisition in a second language? *Psychological Science*, 24(7), 1243–1252. <https://doi.org/10.1177/0956797612472207>
- Gebauer, G. F., & Mackintosh, N. J. (2007). Psychometric intelligence dissociates implicit and explicit learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 33(1), 34–54. <https://doi.org/10.1037/0278-7393.33.1.34>
- Gentry, R. J. (1982). An analysis of developmental analysis in GNYS AT WRK spelling. *The Reading Teacher*, 36(2), 192–200.
- Hayes, H., Treiman, R., & Kessler, B. (2006). Children use vowels to help them spell consonants. *Journal of Experimental Child Psychology*, 94(1), 27–42. <https://doi.org/10.1016/j.jecp.2005.11.001>
- Ise, E., Arnoldi, C. J., Bartling, J., & Schulte-Körne, G. (2012). Implicit learning in children with spelling disability: evidence from artificial grammar learning. *J Neural Transm*, (119), 999–1010. <https://doi.org/10.1007/s00702-012-0830-y>
- Jeffreys, H. (1961). *Theory of probability*. Oxford: UK Oxford University Press.

- Kachergis, G., Yu, C., & Shiffrin, R. M. (2010). Proceedings of the annual meeting of the cognitive science cross-situational statistical learning: Implicit or intentional? *Proceedings of the Annual Meeting of the Cognitive Science Society*, 32(32), 1069–7977.
- Karmiloff-Smith, A. (1992). Nature, nurture and PDP: Preposterous Developmental Postulates? *Connection Science*, 4(3–4), 253–269. <https://doi.org/10.1080/09540099208946618>
- Kaufman, S. B., DeYoung, C. G., Gray, J. R., Jiménez, L., Brown, J., & Mackintosh, N. (2010). Implicit learning as an ability. *Cognition*, 116, 321–340. <https://doi.org/10.1016/j.cognition.2010.05.011>
- Kemper, M. J., Verhoeven, L., & Bosman, A. M. T. (2012). Implicit and explicit instruction of spelling rules. *Learning and Individual Differences*, 22(6), 639–649. <https://doi.org/10.1016/j.lindif.2012.06.008>
- Kessler, B. (2009). Statistical learning of conditional orthographic correspondences. *Writing Systems Research*, 1(1), 19–34. <https://doi.org/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. <https://doi.org/10.1177/0022219412449440>
- Kessler, B., & Treiman, R. (2001). Relationships between sounds and letters in English monosyllables. *Journal of Memory and Language*, 44, 592–617. <https://doi.org/10.1006/jmla.2000.2745>
- Krishnan, S., & Watkins, K. E. (2019). A challenge for the procedural deficit hypothesis: How should we measure sequential learning in childhood? *Developmental Science*, 22(4), 9–10. <https://doi.org/10.1111/desc.12815>

- Lehtonen, A., & Bryant, P. (2005). Doublet challenge: form comes before function in children's understanding of their orthography. *Developmental Science*, 8(3), 211–217.  
<https://doi.org/10.1111/j.1467-7687.2005.00409.x>
- Lelonkiewicz, J. R., Ktori, M., & Crepaldi, D. (2020). Morphemes as letter chunks: Discovering affixes through visual regularities. *Journal of Memory and Language*, 115, 104152.  
<https://doi.org/10.1016/j.jml.2020.104152>
- Marinus, E., Kohnen, S., & McArthur, G. (2013). Australian comparison data for the Test of Word Reading Efficiency (TOWRE). *Australian Journal of Learning Difficulties*, 18(2), 199–212. <https://doi.org/10.1080/19404158.2013.852981>
- Mintz, T. H. (2003). Frequent frames as a cue for grammatical categories in child directed speech. *Cognition*, 90, 91–117. [https://doi.org/10.1016/S0010-0277\(03\)00140-9](https://doi.org/10.1016/S0010-0277(03)00140-9)
- Nigro, L., Jiménez-Fernández, G., Simpson, I. C., & Defior, S. (2015). Implicit learning of written regularities and its relation to literacy acquisition in a shallow orthography. *Journal of Psycholinguistic Research*, 44(5), 571–585. <https://doi.org/10.1007/s10936-014-9303-9>
- Nigro, L., Jiménez-Fernández, G., Simpson, I. C., & Defior, S. (2016). Implicit learning of non-linguistic and linguistic regularities in children with dyslexia. *Annals of Dyslexia*, 66(2), 202–218. <https://doi.org/10.1007/s11881-015-0116-9>
- Nunes, T., Bryant, P., & Olsson, J. (2003). Learning morphological and phonological spelling rules: An intervention study. *Scientific Studies of Reading*, 7(3), 289–307.
- Onishi, K. H., Chambers, K. E., & Fisher, C. (2002). Learning phonotactic constraints from brief auditory experience. *Cognition*, 83, 13–23. [https://doi.org/10.1016/S0010-0277\(01\)00165-2](https://doi.org/10.1016/S0010-0277(01)00165-2)
- 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.

- Pacton, S., Sobaco, A., Fayol, M., & Treiman, R. (2013). How does graphotactic knowledge influence children's learning of new spellings? *Frontiers in Psychology, 4*(OCT). <https://doi.org/10.3389/fpsyg.2013.00701>
- Pavlidou, E. V, Kelly, M. L., & Williams, J. M. (2010). Do children with developmental dyslexia have impairments in implicit learning? *Dyslexia, 16*(1), 143–161. <https://doi.org/10.1002/dys>
- Peirce, J., Gray, J. R., Simpson, S. et al. (2019). PsychoPy2: Experiments in behavior made easy. *Behavior Research Methods, 51*(1), 195–203. <https://doi.org/10.3758/s13428-018-01193-y>
- Pollo, T. C., Treiman, R., & Kessler, B. (2008). Preschoolers use partial letter names to select spellings: Evidence from Portuguese. *Applied Psycholinguistics, 29*, 1–18. <https://doi.org/10.1017/S0142716407080095>
- Powell, M. J. D. (2009). The Bound Optimization BY Quadratic Approximation (BOBYQA) algorithm for bound constrained optimization without derivatives. *Technical Report NA*, 2009/06. Cambridge, England.
- Rastle, K., Lally, C., Davis, M. H., & Taylor, J. S. H. (2021). The dramatic impact of explicit instruction on learning to read in a new writing system. *Psychological Science*. <https://doi.org/https://doi.org/10.1177/0956797620968790>
- Reber, A. S. (1989). Implicit learning and tacit knowledge. *Journal of Experimental Psychology: General, 118*(3), 219–235.
- Qi, Z., Sanchez Araujo, Y., Georgan, W. C., Gabrieli, J. D. E., & Arciuli, J. (2019). Hearing matters more than Seeing : A Cross- Modality Study of Statistical Learning and Reading Ability. *Scientific Studies of Reading, 23*(1), 101–115. <https://doi.org/10.1080/10888438.2018.1485680>

- Reber, A. S., Walkenfeld, F. F., & Hernstadt, R. (1991). Implicit and explicit learning: Individual differences and IQ. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 17(5), 888–896. <https://doi.org/10.1037/0278-7393.17.5.888>
- Reeder, P. A., Newport, E. L., & Aslin, R. N. (2013). From shared contexts to syntactic categories : The role of distributional information in learning linguistic form-classes. *Cognitive Psychology*, 66(1), 30–54. <https://doi.org/10.1016/j.cogpsych.2012.09.001>
- Rosenthal, R. (1979). The file drawer problem and tolerance for null results. *Psychological Bulletin*, 86(3), 638–641. <https://doi.org/10.1037/0033-2909.86.3.638>
- Rouder, J. N. (2014). Optional stopping: No problem for Bayesians. *Psychonomic Bulletin & Review*, 21, 301–308. <https://doi.org/10.3758/s13423-014-0595-4>
- Saffran, J. R., Aslin, R. N., & Newport, E. L. (1996). Statistical learning by 8-month-old infants. *Science*, 274(5294), 1926–1928. <https://doi.org/10.1126/science.274.5294.1926>
- Samara, A., & Caravolas, M. (2014). Statistical learning of novel graphotactic constraints in children and adults. *Journal of Experimental Child Psychology*, 121(1), 137–155. <https://doi.org/10.1016/j.jecp.2013.11.009>
- Samara, A., Singh, D., & Wonnacott, E. (2019). Statistical learning and spelling: Evidence from an incidental learning experiment with children. *Cognition*, 182(March 2018), 25–30. <https://doi.org/10.1016/j.cognition.2018.09.005>
- Samara, A., Smith, K., Brown, H., & Wonnacott, E. (2017). Acquiring variation in an artificial language: Children and adults are sensitive to socially conditioned linguistic variation. *Cognitive Psychology*, 94, 85–114. <https://doi.org/10.1016/j.cogpsych.2017.02.004>
- Schmalz, X., Moll, K., Mulatti, C., & Schulte-Körne, G. (2019). Is statistical learning ability related to reading ability, and if so, why? *Scientific Studies of Reading*, 23(1), 64–76. <https://doi.org/10.1080/10888438.2018.1482304>

- Siegelman, N., Bogaerts, L., & Frost, R. (2017). Measuring individual differences in statistical learning: Current pitfalls and possible solutions. *Behavior Research Methods*, 49(2), 418–432. <https://doi.org/10.3758/s13428-016-0719-z>
- Siegelman, N., Bogaerts, L., Kronenfeld, O., & Frost, R. (2018). Redefining “learning” in statistical learning: What does an online measure reveal about the assimilation of visual regularities? *Cognitive Science*, 42(3), 692–727. <https://doi.org/10.1111/cogs.12556>
- Sobaco, A., Treiman, R., Peereman, R., Borchardt, G., & Pacton, S. (2015). The influence of graphotactic knowledge on adults’ learning of spelling. *Memory and Cognition*, 43(4), 593–604. <https://doi.org/10.3758/s13421-014-0494-y>
- Spencer, M., Kaschak, M. P., Jones, J. L., & Lonigan, C. J. (2015). Statistical learning is related to early literacy-related skills. *Reading and Writing*, 28(4), 467–490. <https://doi.org/10.1007/s11145-014-9533-0>
- Steffler, D. J. (2001). Implicit cognition and spelling development. *Developmental Review*, 21, 168–204. <https://doi.org/10.1006/drev.2000.0517>
- Treiman, R. (1993). *Beginning to spell: A study of first-grade children*. New York: Oxford University Press.
- Treiman, R. (2018). Teaching and learning spelling. *Child Development Perspectives*, 12(4), 235–239. <https://doi.org/10.1111/cdep.12292>
- Treiman, R., & Boland, K. (2017). Graphotactics and spelling: Evidence from consonant doubling. *Journal of Memory and Language*, 92, 254–264. <https://doi.org/10.1016/j.jml.2016.07.001>
- Treiman, R., & Kessler, B. (2015). Choosing between alternative spellings of sounds: The role of context. *Journal of Experimental Psychology: Learning Memory and Cognition*. <https://doi.org/10.1037/xlm0000225>

- Treiman, R., & Kessler, B. (2019). Development of context-sensitive pronunciation in reading: The case of < c > and < g >. *Journal of Experimental Child Psychology*, *182*, 114–125.  
<https://doi.org/10.1016/j.jecp.2019.02.001>
- 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*, *89*(4), e431–e443.  
<https://doi.org/10.1111/cdev.12893>
- Treiman, R., & Wolter, S. (2018). Phonological and graphotactic influences on spellers' decisions about consonant doubling. *Memory and Cognition*, *46*, 614–624.  
<https://doi.org/https://doi.org/10.3758/s13421-018-0793-9>
- Venezky, R. L. (1970). Principles for the design of practical writing systems. *Anthropological Linguistics*, *12*(7), 256–270.
- Vicari, S., Finzi, A., Menghini, D., Marotta, L., Baldi, S., & Petrosini, L. (2015). Do children with developmental dyslexia have an implicit learning deficit? *J Neurol Neurosurg Psychiatry*, *76*, 1392–1397. <https://doi.org/10.1136/jnnp.2004.061093>
- Vidal, Y., Viviani, E., Zoccolan, D., & Crepaldi, D. (2021). A general-purpose mechanism of visual feature association in visual word identification and beyond. *Current Biology*, (31), 1–7.  
<https://doi.org/10.1016/j.cub.2020.12.017>
- Wanzek, J., Vaughn, S., Wexler, J., Swanson, E. A., Edmonds, M., & Kim, A. H. (2006). A synthesis of spelling and reading interventions and their effects on the spelling outcomes of students with LD. *Journal of Learning Disabilities*, *39*(6), 528–543.
- West, G., Vadillo, M. A., Shanks, D. R., & Hulme, C. (2018). The procedural learning deficit hypothesis of language learning disorders: we see some problems. *Developmental Science*, 1–13. <https://doi.org/10.1111/desc.12552>

- Wilkinson, G. S., & Robertson, G. J. (2006). *WRAT4 Wide Range Achievement Test professional manual(4th ed.)*. Lutz, FL: Psychological Assessment Resources.
- Wonnacott, E. (2011). Balancing generalization and lexical conservatism: An artificial language study with child learners. *Journal of Memory and Language*, *65*, 1–14. doi: 10.1016/j.jml.2011.03.001
- Wonnacott, E. (2013). Statistical mechanisms in language acquisition. In P. Binder & K. Smith (Eds.). *The Language Phenomenon* (pp. 65–92). Springer: Berlin, Heidelberg.
- Wonnacott, E., Newport, E. L., & Tanenhaus, M. K. (2008). Acquiring and processing verb argument structure: Distributional learning in a miniature language. *Cognitive Psychology*, *56*(3), 165–209. <https://doi.org/10.1016/j.cogpsych.2007.04.002>
- Yang, J., & Li, P. (2012). Brain networks of explicit and implicit learning. *PLoS ONE*, *7*(8), e42993. <https://doi.org/10.1371/journal.pone.0042993>

**Appendix A. Stimuli used in Experiments 1 and 3**

	list1	list2	list3	list4
exposure	deff	dess	def	des
exposure	dell	dett	del	det
exposure	duf	dus	duff	duss
exposure	dul	dut	dull	dutt
exposure	gess	geff	ges	gef
exposure	gett	gell	get	gel
exposure	gus	guf	guss	guff
exposure	gut	gul	gutt	gull
exposure	meff	mess	mef	mes
exposure	mell	mett	mel	met
exposure	muf	mus	muff	muss
exposure	mul	mut	mull	mutt
exposure	ress	reff	res	ref
exposure	rett	rell	ret	rel
exposure	rus	ruf	russ	ruff
exposure	rut	rul	rutt	rull
legal_unseen	dess	deff	des	def
legal_unseen	dett	dell	det	del
legal_unseen	dus	duf	duss	duff
legal_unseen	dut	dul	dutt	dull
legal_unseen	geff	gess	gef	ges
legal_unseen	gell	gett	gel	get
legal_unseen	guf	gus	guff	guss
legal_unseen	gul	gut	gull	gutt
legal_unseen	mess	meff	mes	mef
legal_unseen	mett	mell	met	mel
legal_unseen	mus	muf	muss	muff
legal_unseen	mut	mul	mutt	mull
legal_unseen	reff	ress	ref	res
legal_unseen	rell	rett	rel	ret
legal_unseen	ruf	rus	ruff	russ
legal_unseen	rul	rut	rull	rutt
illegal	des	def	dess	deff
illegal	det	del	dett	dell
illegal	duss	duff	dus	duf
illegal	dutt	dull	dut	dul
illegal	gef	ges	geff	gess
illegal	gel	get	gell	gett
illegal	guff	guss	guf	gus
illegal	gull	gutt	gul	gut
illegal	mes	mef	mess	meff
illegal	met	mel	mett	mell
illegal	muss	muff	mus	muf
illegal	mutt	mull	mut	mul
illegal	ref	res	reff	ress
illegal	rel	ret	rell	rett
illegal	ruff	russ	ruf	rus
illegal	rull	rutt	rul	rut

**Appendix B. Stimuli used in Experiment 2 (complex constraints)**

	list1	list2	list3	list4
exposure	des	deff	dess	def
exposure	det	dell	dett	del
exposure	duss	duf	dus	duff
exposure	dutt	dul	dut	dull
exposure	geff	ges	gef	gess
exposure	gell	get	gel	gett
exposure	guf	guss	guff	gus
exposure	gul	gutt	gull	gut
exposure	mes	meff	mess	mef
exposure	met	mell	mett	mel
exposure	muss	muf	mus	muff
exposure	mutt	mul	mut	mull
exposure	reff	res	ref	ress
exposure	rell	ret	rel	rett
exposure	ruf	russ	ruff	rus
exposure	rul	rutt	rull	rut
legal_unseen	deff	des	def	dess
legal_unseen	dell	det	del	dett
legal_unseen	duf	duss	duff	dus
legal_unseen	dul	dutt	dull	dut
legal_unseen	ges	geff	gess	gef
legal_unseen	get	gell	gett	gel
legal_unseen	guss	guf	gus	guff
legal_unseen	gutt	gul	gut	gull
legal_unseen	meff	mes	mef	mess
legal_unseen	mell	met	mel	mett
legal_unseen	muf	muss	muff	mus
legal_unseen	mul	mutt	mull	mut
legal_unseen	res	reff	ress	ref
legal_unseen	ret	rell	rett	rel
legal_unseen	russ	ruf	rus	ruff
legal_unseen	rutt	rul	rut	rull
illegal	def	dess	deff	des
illegal	del	dett	dell	det
illegal	duff	dus	duf	duss
illegal	dull	dut	dul	dutt
illegal	gess	gef	ges	geff
illegal	gett	gel	get	gell
illegal	gus	guff	guss	guf
illegal	gut	gull	gutt	gul
illegal	mef	mess	meff	mes
illegal	mel	mett	mell	met
illegal	muff	mus	muf	muss
illegal	mull	mut	mul	mutt
illegal	ress	ref	res	reff
illegal	rett	rel	ret	rell
illegal	rus	ruff	russ	ruf
illegal	rut	rull	rutt	rul