

Running head: IS ENGLISH SPELLING CHAOTIC?

Is English Spelling Chaotic? Misconceptions Concerning its Irregularity

Brett Kessler and Rebecca Treiman

Washington University in St. Louis

## Abstract

An overview of the goals of English orthography counters the misconception that its spelling is chaotic and unprincipled. Direct representation of the speaker's phonemes is not its only goal. But even the sound-to-letter correspondences are not as inconsistent as widely believed. A survey of first-grade text vocabulary shows that spelling consistency is increased significantly if one takes into account the position of the phoneme within the syllable and the identity of the phonemes in the environment. Environmental influences within the rime are especially important. Understanding these patterns may reduce the complexity of spelling for learners and those with spelling problems.

### Is English Spelling Chaotic? Misconceptions Concerning its Irregularity

It is generally believed that the best writing system for a language would be an alphabet that always spells a particular sound in only one way. A person who knew nothing about Finnish except for the sound–letter correspondences could do a credible job of spelling out dictated words or of pronouncing written text. Because of its consistent one-to-one mapping of sounds to letters, Finnish is widely considered to have a nearly optimal orthography. By the same standard, English is generally considered to be “chaotic and indefensible” (Dewey, 1971, p. 4), with the worst orthography of all those that have pretensions to being alphabetic. Dewey's quote, in fact, is one of the kinder remarks made by people who have evaluated the complexity of English spelling. Lists like “*tough, though, through, bough*” have led people to consider English orthography to be hopelessly irregular, a pathological mishmash of correspondences randomly accrued over the past thousand years. With only a touch of irony, G. B. Shaw claimed that “English can't be spelt” (Bett, n.d.). Psychologists and literacy researchers usually express themselves in more sober terms: English has a *deep* orthography (Frost, 1992), or it is *morphophonemic* rather than phonetic. Whatever the nomenclature, English is thought to be far distant from the alphabetic ideal. As a result of these appraisals and of their own observations, some educators have despaired of teaching sound–letter correspondences to beginning spellers. Children are often asked to memorize words as wholes (Scott, 2000), and some people have advocated that the English spelling system be reformed to bring it in line with the ideal of one-to-one mapping (e.g., Dewey, 1971; Lindgren, 1969).

In this paper we wish to state the case for English spelling. We do not want to claim that

the English writing system is ideal, nor do we wish to gloss over the real challenges it poses for children. But it is important to understand the nature of English spelling, and it is seriously misunderstood. English spelling is by no means irrational or pathological, but serves several goals other than that of one-to-one phoneme–letter correspondence that critics have imposed on it. The first part of this paper will briefly describe these additional goals of the spelling system. The second and larger part of this paper will address the issue of regularity. We will argue that English is not nearly as irregular as people think, and will put forward our own ideas about how to measure the regularity, or consistency, of English orthography. In particular, we seek to find out which parts of words have the greatest inconsistency, and to find larger patterns that in effect reduce that inconsistency. Throughout, we focus on the implications of these findings for the teaching of spelling. Early attention to the most productive patterns is expected to make the learning of correct spelling easier and more effective.

### The Principles of English Spelling

Under the most gloomy account of the history of English orthography, the writing system has all but collapsed as a result of centuries of neglect. Like all alphabetic systems, it was meant to have one unique spelling for each of the phonemes in the language. But due in part to the many linguistic and cultural influences on English, there was always a great deal of variation in English spelling, so that even the same word could often be spelled in several different ways. When pressures for standardization arose, a single spelling for each word was chosen in a haphazard fashion from among all the competing spellings, resulting in a huge amount of irregularity in the system. There is more than a grain of truth in that analysis, but it ignores the possibility that a writing system may have goals other than one-to-one mapping between sounds

and letters. In fact, the English spelling system has been shaped by at least three other major principles.

*Conservatism.* Once a spelling is widely accepted, it tends to stick. Although this conservatism is often criticized, it can be useful. If spellings changed, effort would need to be expended to learn the new system; fully literate people would need to learn both systems in order to read recent texts as well as older ones written in the old system. A less obvious benefit of conservatism is that it results in a system that applies equally to many different accents of English. To most speakers of English throughout the world, the distinction between *wh* and *w* is meaningless, because their accents have lost the pronunciation distinction that the spelling distinction originally meant to reflect. But most speakers in Scotland and Ireland still pronounce those two spellings differently. If the spelling system had been less conservative and the spelling had changed as soon as the distinction was lost in London, then what of the speakers in Scotland and Ireland? They would have no motive to adopt a change that would fit their language worse than before. Conservatism serves the function of keeping English spelling reasonably consistent around the globe.

*Unadapted spelling of loan words.* English borrows words freely from other languages, and it almost always uses the spelling of the original language when it does so. Often this unadapted spelling is considered a pretentious nuisance, as if *quiche* is so spelled simply to advertise the fact that one knows French. In reality, the unadapted spelling helps everybody, particularly when the word is first being used in English and is not yet in the English dictionaries. People who know French may immediately recognize the word; those who do not can at least look it up in a French dictionary. If they had encountered the word in the guise *keash*,

they would know how to pronounce it but it would be much harder to ascertain its meaning. Of course, after the word is fully accepted and appears in all the English dictionaries, the situation is a little different. But then the conservatism of English takes over: After the word is fully established in the spelling *quiche*, the spelling sticks. This process of unadapted spelling has a strong impact on the language because very many English words contain Latin and Latinized Greek elements, which are spelled as they were in Latin.

*Representation of nonphonemic information.* The spelling of a word can tell more than just how to pronounce it. Largely as a side-effect of the above two criteria, English now has many sets of words where the spellings tell more about the meanings than the pronunciations alone could. For example, the homophones *sight*, *site*, and *cite* are distinguished, in the first case because the spellings are conservative and represent some sounds whose pronunciations have changed (*sight*, where the *gh* represents a consonant that is no longer pronounced), and also because words from Latin use Latin spelling (the stems of *site* and *cite*). The same principle applies to meaningful parts of words as well (*morphemes*). The reader who encounters the word *citatory* for the first time does not have to waste time considering whether it might have something to do with *site*. This function of morpheme identification has been pointed out by Chomsky (1970). What may be less obvious is that the spelling may tell whether a word is a grammatical (function) word or a lexical (content) word. For example, with rare exceptions, lexical words have at least three letters. The grammatical word *in* is spelled with two letters, but the lexical word *inn* has to be longer. This is achieved by doubling the last letter, resulting in a spelling that looks irregular if one does not understand the principle (Venezky, 1970).

Thus English spelling conveys a good deal more information than just the pronunciation,

and serves other functions as well. Admittedly, for children acquiring literacy, many of these considerations are cold comfort. The writing system does not become any easier for the individual learner from knowing that it applies equally well to multiple dialects and is constant over time. Other features, such as its retention of foreign spellings, only fully benefit people who read the relevant languages. Reminding a six-year child to put a *ph* in *phone* by invoking the Greek word for *sound* would be explaining *obscurum per obscurius*. But the system is not pathological; it is based on principles and does a reasonably good job of applying them.

### The Complexity of English Spelling

Even if we concede that English orthography is not pathologically chaotic, we may still wonder whether it is hopelessly complex, especially for children who are not yet up to speed on their Latin and Greek. To discuss the complexity or consistency of English spelling, it is important to develop a quantitative measure. That not only casts the discussion in objective terms, but it also enables us to convincingly compare the relative complexity of the different parts of words.

Attempts at measuring the complexity of English spelling usually begin, and end, by observing how many different spellings a given sound has in different words. For example, Dewey (1971; Appendix A) lists all the different spellings for each phoneme. The entry for /ε/ (see International Phonetic Association, 1996, 1999 for a description of the phonetic symbols used in this paper) lists *many* (*a*), *said* (*ai*), *says* (*ay*), *men* (*e*), *ledge* (*e\_e*), *head* (*ea*), *cleanse* (*ea\_e*), *keelson* (*ee*), *belles-lettres* (*e\_e*), *heifer* (*ei*), *leopard* (*eo*), *cheque* (*e\_ue*), *friend* (*ie*), *bury* (*u*), and *guess* (*ue*). One way people have quantified such displays is by simply counting how many different spellings are found per phoneme. A typical figure is that there are about twelve

spellings per phoneme; or, taking the ratio, English has a consistency of .08 (8%) (Hotson, as cited in Dewey, 1971). Such figures are of a piece with claims like that of Dewey, who calculated that the word *taken* could be spelled 5,157,936 different ways: These computations take into account only the number of different spellings per sound regardless of where it is found in the word. The implicit psychological model of the spelling process in such formulations is that the speller sounds out the word phoneme by phoneme; for each phoneme recalls at random one of the possible spellings for that phoneme; writes it down and moves on to the next phoneme. With such a method, virtually no word would end up being spelled correctly. The implication, therefore, is that spellers instead have to memorize the spelling of each word as a whole (Lindgren, 1969).

Hanna, Hanna, Hodges, and Rudorf (1966), in the first large-scale computer-assisted study of English spelling, discussed a somewhat less pessimistic model of spelling. Why would children bother to learn a dozen spellings for each sound if applying them at random is virtually guaranteed to be wrong? The most logical strategy would be to just memorize for each phoneme the spelling that is used in the largest number of words, and ignore the rest. On this model, Hanna et al. calculated, 73% of all phonemes in text would be spelled correctly. Of course, most words have more than one phoneme, and in the end, most words would still be spelled incorrectly. Nevertheless, quite a few words would be spelled correctly (40% of the words in text, according to Dewey, 1971), significantly reducing the number of words that would have to be memorized whole.

There are, however, a number of reasons to think that even Hanna et al.'s (1966) figure is a rather low estimate of the consistency of English spelling, especially if our main concern is the



difficulty it imposes on children.

*Vocabulary level.* Many sound-to-spelling correspondences in English do not turn up in the vocabulary of young children. It may be years before people really need to learn that /ɛ/ is spelled *ee* in *keelson*. If we do not require our spelling model to handle such cases, we will not have to count them as inconsistent spellings. Therefore it is important to consider only words that would be reasonably familiar to young children.

*Position.* One supposes that Shaw was being deliberately provocative when he claimed (as cited, e.g., in Vachek, 1973) that *ghoti* would be a perfectly regular (i.e., rule-governed) spelling for the word /fɪʃ/ (*fish*). But one finds this claim repeated so often that one gets the impression that many people do not perceive the irony. In fact, *gh* never spells /f/ at the beginning of a word; *ti* never spells /ʃ/ at the end of a word. Position is a very important factor in English spelling, and is one that children learn readily and easily. For example, children learn very early that double letters do not belong in word-initial position (Treiman, 1993).

*Environment.* In many cases, it is obvious that the spelling for a particular sound can be influenced, or *conditioned*, by other elements in the word. The popular mnemonic “*i* before *e* except after *c*” capitalizes on environment: It states that *ei* as a spelling for /i/ is more common after *c* than after other letters. The role of environment in spelling has not been studied very extensively, however. Scholars such as Cummings (1988) have discussed the matter in a qualitative way, and Venezky (1970, 1999) has done the same for reading, but they did not give explicit measures or specifically address children’s vocabulary.

Not paying attention to environment can lead to estimates of the consistency in English spelling which are much too low. Consider, for example, the vowel /a/ in the General American

accent. In most words, it is spelled *o*, as in *rock*, *top*, *pot*, and so forth. But when it is followed by the consonant /r/ in the same syllable, the same vowel is spelled *a*, as in *car*, *start*, *harp*, and so forth. If children were completely insensitive to environment, one might think that all the dozens of words like *car* have to be treated as exceptions, and perhaps memorized whole. But if children learn one simple rule of environmental conditioning, all of those *ar* spellings become completely regular.

*Common semantics.* As we discussed earlier, much of the inconsistency in sound-to-spelling rules is due to the fact that English spelling conveys more information than just pronunciation. Spelling becomes easier to the extent that a speller is sensitive to those additional factors. In particular, once a child has learned a word containing a morpheme with an irregular spelling, then other words containing the same morpheme are usually much easier to spell. For example, *foot* is arguably an inconsistent spelling of General American /fʊt/: Why is it not spelled *fut*, parallel to *put*? But once that word is learned, *footing*, *football*, *footpath*, and so forth, come almost for free. It would be misleading to count those words when measuring total inconsistency in the language.

*Parallel processing and statistical learning.* While the issues considered so far have to do with properties of the spelling system itself, other important issues deal with the nature of human learning and cognitive performance. It is, perhaps, easiest to imagine spelling as a straightforward process where spellers do one operation at a time, in a strictly defined sequence that is designed to be of maximum efficiency. For example, perhaps they spell one phoneme at a time, generating the most common spelling of each phoneme. The whole spelling process could be followed by a visual check, and if the word is not recognized, then the speller will attempt to

retrieve the spelling, as a whole, from memory. Under such a model, the speller would pay attention to the spellings of individual phonemes, but would have no reason to remember minority spellings (such as *c* for initial /s/ as in *civil*) once it was confirmed that some other spelling (*s*) is the most common. Nor would a speller have any reason to pay attention to how the spelling of one sound is conditioned by other sounds in its environment.

Such models are logical, efficient, and easy to describe. But it would be rather surprising if children naturally learned to spell that way. It would be as if a child, learning to identify animals, quickly decided that they were most efficiently identified by their most salient characteristic, perhaps their call, and would not bother to learn any other properties of the animals. To identify an animal the child would only attend to its call; if a dog refused to bark, the child would, in this scenario, be entirely mystified as to its identity. It would be very easy to model this animal recognition device, and its only flaw is that it is clearly wrong. Children attend to many different properties of the animals they learn to identify, and on encountering a new individual, they process several different pieces of information. Furthermore, this processing is not done in serial way, as in a flowchart. Rather, the various pieces of information are processed in parallel.

It is reasonable to assume that a good reader and speller would approach the English writing system in the same way. We would not expect that children would necessarily learn and use only the most frequent spelling for a particular pronunciation. Rather, we would expect that they would to some extent learn all of the spellings they encounter, although they would eventually gain some sense of the relative frequency of the various spellings. Minority spellings can be exploited because learning is often imperfect. When a child encounters the word *sigh*, she

or he may simply remember that there was something uncommon about the spelling, or perhaps that there was an *h* or some silent consonant. Coupled with the knowledge that *igh* is a minority representation for /aɪ/, this may be enough information to reconstruct the correct spelling of the word.

Because of the problems with previous analyses of spelling consistency and the spelling models on which they were based, we developed a new measure of consistency. The following section describes how we derived that measure and applied it to our own analysis of the complexity of English spelling (Kessler & Treiman, 2001), while avoiding the pitfalls that were just discussed. In particular, far from ignoring position and environment, our analysis will focus on the roles they can play in facilitating children's spelling.

#### An Analysis of English Spelling Consistency

*Selecting the data.* The problem of vocabulary level was addressed by consulting the word list of Zeno, Ivens, Millard, and Duvvuri (1995), which tells how many times words appeared in a large sample of reading material used in schools in the United States. We selected only words for which Zeno et al. reported an adjusted frequency value of at least 20 words per million for kindergarten and first-grade texts. The problem of common semantics was addressed by excluding words that share their root with some simpler word. For example, *ninth* was not included in our analysis because it is an extension of *nine*, which is included. We used only one-syllable words in this study. Our primary motivation for this restriction was simplicity. If we are to take into consideration the issues of position and environment, the analysis is difficult enough to carry out and to understand when we are dealing with the various parts of a one-syllable word and their effects on each other. It is best to first establish the methodology, and

later extend it to more complex cases. The results we report will therefore be most appropriate to educational environments where pupils first begin studying one-syllable words and only later attempt to spell multiple-syllable words. After we rejected words that had multiple syllables or shared roots and did not appear often in texts intended for kindergartners and first-graders, we had 914 words. All of the words were listed in their most common U.S. spelling and General American pronunciation.

*Treatment of position and environment.* There are many different ways to divide words up in order to take position and environment into account. At first, one might consider categorizing sounds and letters by their absolute position in the word, e.g., the second from the start or the third from the end. But it is not very likely that the *t* in *star* and the *a* in *tank* would behave similarly just because they are both the second letter from the start. Instead, we adopted an approach based on phonological structure, centering the analysis about the vowel. The vowel is the nucleus of the syllable, and the one phoneme type that is mandatory. Vowels are also notoriously difficult to spell in English (e.g., Treiman, 1993), so from a pragmatic point of view, it is very useful to be sure that we treat the vowel specially. So we divided the word into three parts: the vowel; the optional consonants that precede the vowel, called the *onset*; and the optional consonants that follow the vowel, called the *coda*. This division allowed us to confirm whether the vowel is indeed the least consistent part of the word. Further, it allowed us to ask whether environment makes the vowel any easier to spell, and, if so, which of the two other parts of the syllable helps the most: the onset or the coda. At the same time, we could investigate other relationships, such as whether knowing the vowel helps us to spell the onset or the coda.

After dividing the phonemes of each word into onset, vowel, and coda, the next step was

to decide which letters in the standard spelling correspond to which of those three parts of the syllable. We assigned all letters to one of those three parts; no letters were left out on the grounds that they were silent. Table 1 gives several examples of how the spellings were divided. In many words it was not clear whether certain letters were better analyzed as spelling the vowel or the coda; in such cases, we assigned the letters to the vowel, e.g., *c-augh-t* and *t-al-k*. We assigned final Silent E to the vowel or the coda depending on its function. To be specific, when the vowel was otherwise spelled with a single vowel letter and followed by no more than one consonant spelled with a single letter, or by /st/, the Silent E was assigned to the vowel. We assigned the E to the coda when the letter before the Silent E was *c*, *g*, *s*, *z*, *u*, *v*, or *th*. An example of coda assignment is *prince*, where the Silent E does not make the vowel long, but indicates that the last *c* is pronounced /s/ rather than /k/. In many words, such as *prize*, both of these rules applied, and so the E was assigned both to the vowel and to the coda.

*Consistency measure.* Once the spellings and pronunciations of all the words were divided into onset, vowel, and coda, we were in a position to measure the spelling consistency of each of those syllable positions. We started by measuring the consistency of particular sounds in particular positions. Table 2 gives, by way of example, all of the words where the sound in the vowel position is /aɪ/. For example, the table shows that the spelling *i\_e* is used in .594 of the words (41/69). We expressed the spelling consistency of a particular vowel by taking the weighted average of these proportions across each of its spellings. That is, we multiplied each proportion by the number of words it accounts for; added those products together; then divided by the total number of words. In this case, the vowel /aɪ/, the result was .398. Our consistency measure is exactly 1 when the spelling of a sound is perfectly consistent (i.e., only one spelling),

and gets smaller the more distinct spellings there are and the more evenly spread the frequencies of those spellings are. For example, if a sound can be spelled two different ways and appears in 100 different words, the measure will be .500 if the two spellings are equally frequent, but will be .980 if one of the two spellings appears only in a single word. This difference reflects the fact that the spelling is more unpredictable in the former case. The more inconsistent the spelling is, the closer the measure approaches 0.

By this procedure, we obtained consistency measures for each of the vowels. To get an overall consistency measure for vowels in general, we took the weighted average of those consistency measures, a procedure that counts more heavily the consistency of vowels that appear in the largest number of words. This gave a consistency measure of .509 for the vowels. Following the same procedure for the onsets yielded the number .921; for the codas, .854 (Table 3, first row). If we wish to combine the figures for the three positions, we get an overall average of .761. That is equivalent to the complexity of a system where every sound has two spellings, with one of those spellings being used 86% of the time. Such numbers could be used to quantify precisely how much the consistency of English spelling differs from that of other languages. These measurements are clearly a good deal higher than the extreme figures offered by some critics of English spelling, such as the 8% given by Hotson (as cited in Dewey, 1971). Furthermore, these numbers take into account the role of position. We now have clear quantification of how much harder it is to spell vowels than consonants. Moreover, we see that consonants are easier to spell in onset position than in coda position. This ranking of consistencies agrees with findings that children on average spell onsets better than codas, and codas better than vowels (Treiman, Berch, & Weatherston, 1993). The implications are clear for

educators who may wish to emphasize the less frustrating tasks in the earlier stages of learning, or to allot more time in later stages to perfect the more difficult components of spelling.

*Effects of environment.* Next we addressed the question of how the environment can help spelling. For example, when a child attempts to spell a vowel, is the task simpler if the consonants in the word are taken into account? If so, which consonants help more, those of the onset or of the coda? These questions can be answered by calculating *conditional consistencies*. For example, we might start by computing the consistency of a particular vowel, such as /aɪ/, in all words that start with a particular onset, say /b/: That would be the conditional consistency of /aɪ/, given /b/ as the onset. If we do that same calculation for each of the different vowels and take the weighted average of the answer, we get the conditional consistency of vowels in general, given /b/ as the onset. Lastly, if we do that same computation—finding the conditional consistency of vowels given a particular onset—for all possible onsets, and take the weighted average of all those conditional consistencies, we end up with what can be described simply as the conditional consistency of vowels given the onset. That figure is .721, which is a good deal higher than the unconditional consistency of .509. We did the same sort of computation six times: the conditional consistency of each of the three syllable positions, given each of the other two syllable positions. The results of these computations are presented in Table 3. The top row gives the unconditional consistency of the syllable position named in the column header, and subsequent rows show the consistency of that position when it is conditioned by the position named in the row header. These data make clear that considering the environment helps spelling, and often by an appreciable amount. When confronted with inconsistent syllable parts, especially the vowel, the child does not necessarily need to fall back on memorized spellings of the entire



word.

The fact that the conditional consistencies are all higher than the unconditional consistencies is not very surprising. Mathematically, a conditional consistency can never be lower than an unconditional consistency, and there are good reasons for expecting it to be higher in this particular task. Consider Table 2 again. It will be noticed that /aɪ/ is always spelled *i\_e* after /d/ (i.e., *die* and *dime*). The conditional consistency is a perfect 1.000 after that onset. To our knowledge, there is no inherent reason why that spelling is always used in that environment, other than pure coincidence. There are many different onsets, but relatively few words containing /aɪ/. Therefore it is unavoidable that some onsets appear with some spellings in numbers quite a bit larger than the average. This makes the conditional consistency go up, but the cause of those elevated sound–spelling associations is pure coincidence.

Are the increases in consistency that we got by considering environment solely attributable to those kinds of coincidences? We pursued this question in a very direct way. If the rise in vowel consistency from .509 to .721 when the onset is considered is due to chance, that is the same as saying that if the onsets were randomly switched among all the words, the conditional consistency would still typically be in the neighborhood of .721. So we did just that, mixing the onsets randomly across all the words 10,000 times and counting how many of those 10,000 rearrangements had a conditional consistency of at least .721. It turned out that 35% of all the rearrangements had such a high conditional consistency. That percentage, by definition, is the significance value, *p*, of our test, and it is far from attaining the .05 value generally required in psychology. Thus, we have not proved that there is any principled reason why onsets help predict the spelling of vowels. Similar tests showed that any improvements between the two consonant

positions—the consistency of onsets given codas, or of codas given onsets—are also coincidental in the same sense. In contrast, for the other three conditional probabilities, marked with an asterisk in Table 3, random rearrangements of the conditioning syllable position yielded consistencies that matched the original no more than one time in a thousand. That is, the rise in vowel consistency when one considers the coda, and the rise in consonant consistencies (onset or coda) when one considers the vowel, are not coincidental.

We should not attribute too sweeping a role to this distinction between coincidental and noncoincidental rises in consistency. Humans can and do learn patterns that are coincidental; perhaps at some level some children do profit from learning some coincidental generalizations like “/aɪ/ is spelled *i\_e* after /d/.” What these figures do tell us is that children stand to profit a good deal more from paying attention to some environments than others. When spelling the onset, accuracy can be improved significantly if the child considers the vowel; any improvement gained from considering the coda is coincidental and of smaller magnitude. When spelling the vowel, the larger and more significant improvement comes from considering the coda. When spelling the coda, the larger and more significant improvement comes from considering the vowel.

Most of the improvements to be gained from considering environment take place between the vowel and the coda. For Figure 1, we have calculated by how much the conditional consistencies are higher than the conditional consistencies one would expect by chance (measured as the average conditional consistencies across all the 10,000 rearrangements). For example, the figure shows that the vowel spelling becomes 14.3% more predictable when the coda is taken into consideration, over and above chance effects. It is noteworthy that influences

between vowel and coda go in both directions. In contrast, the influence between vowel and onset is unidirectional (the onset does not help spell the vowel above chance levels) and much smaller: The consistency of the onset goes up only 1.2% over chance when the vowel is taken into account, as compared to increases of from 7.5% to 14.3% between vowel and coda. These statistics are in line with a great deal of research that shows that the vowel and coda form a special phonological domain, called the *rime*, and that the connections between elements within that domain are much stronger than any connection between the onset and vowel. Not only is this domain important for linguistic descriptions of languages like English, but it is also the case that children more readily treat the vowel and coda as a unit than they do the onset and vowel (for a summary of such research, see Treiman & Kessler, 1995). For example, when asked to divide a syllable in two, children find it easier to break it after the onset, keeping the vowel and coda intact.

These findings suggest that a strong emphasis should be put on rimes when teaching spelling. Not only are children inherently more capable of dealing with rimes as a unit, but that is precisely the most profitable strategy when spelling English words. It may be helpful to generally present spelling words in groups like *child, wild, mild* vs. *side, ride, hide*. Not only would that approach be most likely to turn up statistically important patterns in specific cases (as in this example), but children would also profit just from internalizing the strategy of considering riming words when learning and recalling spellings. To a lesser extent, the findings also suggest that onsets that have multiple spellings, such as /k/, may sometimes be best taught in sets of words that are arranged by the following vowel (e.g., *can, cat, catch* vs. *kid, kiss, kit*). To be sure, children need to be prepared for the fact that several patterns have exceptions; not all rimes, for

example, are spelled identically in all words (e.g., *right* vs. *white*). But children can still make use of patterns, even if they have exceptions. In any event, the number of exceptions is lower than would be encountered if children considered only single phonemes individually without regard for environment.

*Specific environment patterns.* Up to this point we have been discussing patterns between onsets, vowels, and codas in the abstract. In addition to those general tests, we also used statistical tests to determine which individual sounds are significantly more consistent when environment is considered. We ran these tests for the three greater-than-chance relations found in Figure 1: We asked which onsets are helped by which vowels, which vowels are helped by which codas, and which codas are helped by which vowels.

Turning our attention first to the rime, where the biggest effects are located, we found that 4 coda types are improved by considering the vowel, using the cutoff of .05 for statistical significance. When /l/ or /s/ is alone in the coda, it tends to be spelled with a double letter when the vowel is /æ/ (*shall, class*), /ɑ/ (*doll*), /ʌ/ (*dull, fuss*), /ɛ/ (*bell, dress*), /ɪ/ (*fill, kiss*), /ɔ/ (*ball, boss*), or /ʊ/ (*pull*), but with a single consonant letter after other vowels (e.g., *jail, goose*). The coda *k* shows an alternation between *ck* after /æ/ (*back*), /ɑ/ (*block*), /ʌ/ (*duck*), /ɛ/ (*neck*), and /ɪ/ (*chick*), versus a single *k* elsewhere (e.g., *cake, walk*). What these patterns all have in common is that they use two consonant letters when the vowel is spelled with a single letter, and a shorter spelling when the vowel is spelled with more than one letter, counting Final E in the tally. The fourth coda, /z/, also has a special spelling (*s*) after the same set of vowels that condition *ll*, *ss*, and *ck* (e.g., *as, is, was*).

Figure 1 shows that the reverse influence, that of codas on vowel spelling, is stronger

than that of vowels on coda spelling. Correspondingly, we found that there are many vowels that can be spelled significantly more accurately when the coda is considered: 13 of the 15 vowel types. Table 4 lists most of them, along with the spellings that become significantly more common when the vowel is followed by certain codas. For example, when the vowel /aɪ/ appears before the coda /t/, the spelling *igh* becomes much more frequent than it is when not before /t/. This table corroborates the general finding that codas strongly help predict vowel spelling. Moreover, it shows that the patterning is very pervasive, and gives an idea of what patterns may be most helpful to emphasize in guiding beginning spellers and those who have experienced difficulties.

Finally, turning our attention to the rather small influence that crosses the onset–rime boundary (the 1.2% above-chance influence of the vowel on the onset spelling), it is not surprising that only one phoneme is improved significantly: /k/ is significantly more likely to be spelled *k*, as opposed to *c*, before the vowels /i/ (*keep*, *key*), /ɪ/ (*kick*, *kill*), and /aɪ/ (*kind*, *kite*). (An even more useful version of this pattern is that /k/ is spelled *k* before the letters *e* and *i*, but we did not specifically study the effects of spelling on spelling; it may be difficult for children to learn to take into account the effects of letters that they have not yet written down.) There are, to be sure, other instances where the vowel conditions the onset spelling in English, but none of them are very prominent in the child vocabulary we have considered. It may be better for children to concentrate on learning patterns that they will see reinforced repeatedly in their reading and writing, such as this effect of the vowel on the spelling of /k/, than on patterns that will not become particularly salient until they are older. It may seem disappointing that environment is of relatively small help in spelling onsets, but it should be kept in mind that

onsets are already highly consistent and easy to spell. It is precisely where help is most needed—in spelling the less consistent codas and especially the vowels—that environment is most helpful.

One might fear that the patterns identified in the written vocabulary of young children would be peculiar to that vocabulary and no longer apply when more words are learned. Fortunately, it turns out that virtually all of the patterns hold for adult vocabulary as well (Kessler & Treiman, 2001). Therefore if a child learns patterns such as “*igh* before /t/”, whether explicitly or implicitly by seeing lists such as *night*, *light*, *right*, etc., that should not only help with current vocabulary but also serve the child for life.

We might note in passing that our conclusions apply in broad measure to reading as well as to spelling. Reading is more consistent than spelling, and some of the details vary because consistency is not symmetrical. For example, the spelling of the onset sound /f/ is not completely consistent because sometimes it is spelled *f* and sometimes *ph*; but the readings of the onset letters *f* and *ph* are completely consistent, because both are always pronounced /f/. Despite these differences, it turns out that, as in spelling, some syllable positions become significantly more consistent when environment is taken into account, and the strongest effects are within the rime. Details can be found in such works as Kessler and Treiman (2001) and Stanback (1992).

Our discussion to this point has been motivated by general learning theory. Humans are pattern learners, and so it makes sense that they might pick up on the sort of patterns we have described. But is there any direct evidence that they do so? Treiman, Kessler, and Bick (2002) tested whether college students are sensitive to onset and coda environments when spelling vowels. They asked the students to spell nonsense words such as /glart/ and /glarb/. If spellers

were not sensitive to environment, they would spell /aɪ/ the same way in both words. But in fact, they use the *igh* spelling more often when spelling nonsense words like /glart/, which end in /t/, than when spelling nonsense words that end in other consonants, as does /glarb/. This demonstrates that adult spellers do not blindly spell phoneme by phoneme, but that they take into account how environment affects spelling: As Table 4 shows (row 5), the coda /t/ is a strong conditioner of the spelling *igh* for the vowel /aɪ/. Another experiment showed that when college students misspell words, the errors tend in the direction of overgeneralizing these regular patterns. For example, spellers mistakenly used the vowel spelling *ea* in words like *shred* more often than they did in words like *fleck*, showing sensitivity to the pattern that *ea* is an especially common spelling of /ɛ/ when the coda is /d/. These effects were stronger among better spellers. One interpretation of this finding is that better spellers are more sensitive to environmental patterns than are poorer spellers. At this point it is premature to say that cause and effect have been conclusively demonstrated, and we are still in the process of performing similar studies on young children. But it is not too much of a stretch to infer that poor spellers may be given a leg up by calling their attention to environmentally conditioned spelling patterns.

### Conclusion

English is not Finnish. Its spelling system requires years of study to master, and many pupils find it frustrating. But the widespread belief that it is chaotic and unprincipled arises from a misconception that its only goal is to express the sounds of the speaker's accent. Once we understand its additional goals, whether or not we personally agree with them, it is easier to see that English generally follows them in a principled way. And even if we do restrict our purview to sound–letter correspondences, common misconceptions about their degree of inconsistency

can be overcome by using measures that do not assign undue importance to rare spellings, and by considering the effects of position and environment. The knowledge that English is more principled and consistent than commonly believed should help in teaching spelling to normally developing children as well as to those who find spelling problematic.



## References

- Bett, S. T. (n.d.). The trouble with spelling: The nature of and alternatives to standard English orthography. Retrieved January 6, 2003, from <http://victorian.fortunecity.com/vangogh/555/Spell/trublsp11.html>.
- Chomsky, C. (1970). Reading, writing, and phonology. *Harvard Educational Review*, 40, 287–309.
- Cummings, D. W. (1988). *American English spelling: An informal description*. Baltimore: Johns Hopkins University Press.
- Dewey, G. (1971). *English spelling: Roadblock to reading*. New York: Teachers College Press.
- Frost, R. (1992). Orthography and phonology: The psychological reality of orthographic depth. In M. Noonan, P. Downing, & S. Lima (Eds.), *The linguistics of literacy* (pp. 255–274). Amsterdam: Benjamins.
- Hanna, P. R., Hanna, Jean S., Hodges, R. E., & Rudorf, E. H., Jr. (1966). *Phoneme–grapheme correspondences as cues to spelling improvement*. Washington, DC: U. S. Department of Health, Education, and Welfare.
- International Phonetic Association. (1996). Reproduction of The International Phonetic Alphabet. Retrieved from <http://www2.arts.gla.ac.uk/IPA/ipachart.html>.
- International Phonetic Association. (1999). *Handbook of the International Phonetic Association: A guide to the use of the International Phonetic Alphabet*. Cambridge, Eng.: Cambridge University Press.
- Kessler, B., & Treiman, R. (2001). Relationships between sounds and letters in English monosyllables. *Journal of Memory and Language*, 44, 592–617.

- Lindgren, H. (1969). *Spelling reform: A new approach*. Sydney, Australia: Alpha.
- Scott, C. M. (2000). Principles and methods of spelling instruction: Applications for poor spellers. *Topics in Language Disorders, 20*, 66–82.
- Stanback, M. L. (1992). Syllable and rime patterns for teaching reading: Analysis of a frequency-based vocabulary of 17,602 words. *Annals of Dyslexia, 42*, 196–221.
- Treiman, R. (1993). *Beginning to spell*. Oxford: Oxford University Press.
- Treiman, R., Berch, D., & Weatherston, S. (1993). Children's use of phoneme–grapheme correspondences in spelling: Roles of position and stress. *Journal of Educational Psychology, 85*, 466–477.
- Treiman, R., & Kessler, B. (1995). In defense of an onset-rhyme syllable structure for English. *Language and Speech, 38*, 127–142.
- Treiman, R., Kessler, B., & Bick, S. (2002). Context sensitivity in the spelling of English vowels. *Journal of Memory and Language, 47*, 448–468.
- Vachek, J. (1973). *Written language: General problems and problems of English*. The Hague: Mouton.
- Venezky, R. L. (1970). *The structure of English orthography*. The Hague: Mouton.
- Venezky, R. L. (1999). *The American way of spelling: The structure and origins of American English orthography*. New York: Guilford.
- Zeno, S. M., Ivens, S. H., Millard, R. T., & Duvvuri, R. (1995). *The educator's word frequency guide*. Brewster, NY: Touchstone Applied Science Associates.

Author Note

Brett Kessler and Rebecca Treiman, Department of Psychology, Washington University in St. Louis.

This work was supported by National Science Foundation Grants SBR-9807736 and BCS-0130763. Correspondence concerning this article should be addressed to Brett Kessler, Psychology Department, Washington University in St. Louis, Campus Box 1125, One Brookings Drive, St. Louis MO 63130-4899, USA. Email: [bkessler@BrettKessler.com](mailto:bkessler@BrettKessler.com).

Table 1

*Sample Divisions of Words by Part of Syllable*

Word	Onset	Vowel	Rime
<i>talk</i>	t	al	k
	/t/	/ɔ/	/k/
<i>name</i>	n	a_e	m
	/n/	/e/	/m/
<i>taste</i>	t	a_e	st
	/t/	/e/	/st/
<i>prince</i>	pr	i	nce
	/pr/	/l/	/ns/
<i>course</i>	c	ou	rse
	/k/	/ɔ/	/rs/
<i>prize</i>	pr	i_e	ze
	/pr/	/aɪ/	/z/

Table 2

*Data for Computing the Consistency of a Vowel (/aɪ/)*

Spelling	Words	Count	Proportion
i_e	bike, bite, die, dime, drive, fine, fire, five, hide, ice, kite, knife, lie, life, like, line, live, mice, mike, mine, nice, nine, pie, pile, pine, pipe, prize, quite, ride, shine, size, smile, tie, time, twice, while, white, wide, wife, wise, write	41	.594
igh	bright, fight, high, light, might, night, right, sight, tight	9	.130
y	by, cry, dry, fly, guy, my, sky, try, why	9	.130
i	child, climb, find, kind, mind, wild, wind	7	.101
uy	buy	1	.014
eye	eye	1	.014
ig	sign	1	.014
SUM		69	

Table 3

*Spelling Consistencies of Each Syllable Position*

Given	Onset	Vowel	Coda
(Unconditional)	.921	.509	.854
Onset	–	.721	.936
Vowel	.958 <sup>*</sup>	–	.970 <sup>*</sup>
Coda	.930	.797 <sup>*</sup>	–

\* $p \leq .001$ .

Table 4

*Vowel Spellings Conditioned by Codas*

Vowel	Coda	Spelling	Examples
/aɪ/	General	<i>i_e</i>	<i>like, time, side, nice, tie, lie, pie</i>
	None	<i>y</i>	<i>my, why, by, sky, fly, try, cry, dry</i>
	/ld/	<i>i</i>	<i>child, wild</i>
	/nd/	<i>i</i>	<i>find, mind, kind</i>
	/t/	<i>igh</i>	<i>right, night, light, might, bright, fight, tight, sight</i>
/aʊ/	General	<i>ou</i>	<i>out, house, mouth, south, shout</i>
	None	<i>ow</i>	<i>now, how, cow</i>
	/n/	<i>ow</i>	<i>down, brown, town, clown</i>
/ɑ/	General	<i>o</i>	<i>not, got, stop, box, hot, lot, top, job, shop, rock</i>
	/r/ ± consonant	<i>a</i>	<i>car, far, jar, bar, hard, dark, mark, park, start, part, arm, farm, yard, smart, large, start, bark, card</i>
	/ʌ/	<i>u</i>	<i>but, up, us, must, run, much, fun, sun, lunch, jump</i>
/ɒ/	/m/	<i>o_e</i>	<i>some, come</i>
	/v/	<i>o_e</i>	<i>love, dove</i>
/e/	General	<i>a_e</i>	<i>came, make, take, made, name, gave, same, face</i>
	None	<i>ay</i>	<i>day, way, play, say, may, stay, pay, gray, lay, ray</i>
	/l/	<i>ai</i>	<i>tail, mail, sail, pail, trail, nail, jail</i>

Vowel	Coda	Spelling	Examples
	/n/	<i>ai</i>	<i>rain, train, pain, main</i>
/ɛ/	General	<i>e</i>	<i>then, get, went, when, them, help, tell, well, yes</i>
	/d/	<i>ea</i>	<i>head, bread, dead, thread</i>
	/r/	<i>a_e</i>	<i>care, share, scare</i>
	/r/	<i>ai</i>	<i>air, hair, fair, chair, pair</i>
	/r/	<i>e_e</i>	<i>there, where</i>
/i/	General	<i>ea</i>	<i>eat, each, mean, please, clean, leave, seat, cream</i>
	None	<i>e</i>	<i>he, she, we, me, be</i>
	/d/	<i>ee</i>	<i>need, feed, speed, weed</i>
	/p/	<i>ee</i>	<i>keep, sleep, sheep, deep</i>
/ɪ/	General	<i>i</i>	<i>it, in, is, his, will, with, this, did, big, him, if, fish</i>
	/r/	<i>ea</i>	<i>hear, near, year, dear, ear</i>
	/v/	<i>i_e</i>	<i>give, live</i>
/o/	General	<i>o_e</i>	<i>home, those, hole, close, nose, hope, rope, note</i>
	None	<i>ough</i>	<i>though, dough</i>
	None	<i>ow</i>	<i>show, grow, snow, blow, slow, throw, low, crow</i>
	/d/	<i>oa</i>	<i>toad, road</i>
	/ld/	<i>o</i>	<i>old, told, cold, hold, gold, sold, fold</i>
	/st/	<i>o</i>	<i>most, ghost</i>



Vowel	Coda	Spelling	Examples
	/t/	<i>oa</i>	<i>boat, coat, goat</i>
/ɔ/	General	<i>o</i>	<i>dog, long, off, lost, wrong, strong, boss, soft, cloth</i>
	None	<i>aw</i>	<i>saw, draw, straw, paw</i>
	/k/	<i>al</i>	<i>walk, talk</i>
	/l/	<i>a</i>	<i>all, ball, small, call, fall, tall, hall, wall</i>
	/r/	<i>o_e</i>	<i>more, store, shore, wore</i>
	/r/	<i>oo</i>	<i>floor, door</i>
	/t/	<i>ough</i>	<i>thought, brought, bought</i>
/u/	General	<i>oo</i>	<i>school, room, soon, food, moon, pool, tooth, cool</i>
	None	<i>ew</i>	<i>new, few, blew, grew, flew, threw, drew</i>
	None	<i>o</i>	<i>to, do, who</i>
	/p/	<i>ou</i>	<i>soup, group</i>
	/t/	<i>ui</i>	<i>suit, fruit</i>
/ʊ/	General	<i>oo</i>	<i>look, took, cook, foot, shook, poor, book, hook</i>
	/d/	<i>oul</i>	<i>would, should, could</i>
	/l/	<i>u</i>	<i>pull, full, bull</i>
	/ʃ/	<i>u</i>	<i>push, bush</i>

*Note.* “General” coda environments are those other than the ones listed. Examples are in

descending frequency in child vocabulary.

Figure Captions

*Figure 1.* Amount by which the phonemes in one syllable position raise the consistency of the spelling of another syllable position (that pointed to by arrow), over chance levels.

