Published in final edited form as:

Sci Stud Read. 2016 ; 20(5): 349–362. doi:10.1080/10888438.2016.1186168.

Measures of Kindergarten Spelling and Their Relations to Later Spelling Performance

Rebecca Treiman,

Washington University in St. Louis

Brett Kessler,

Washington University in St. Louis

Tatiana Cury Pollo,

Universidade Federal de São João del-Rei

Brian Byrne, and

University of New England and Australian Research Council Centre of Excellence in Cognition and its Disorders; National Health and Medical Research Council Centre of Excellence in Twin Research

Richard K. Olson

University of Colorado

Abstract

Learning the orthographic forms of words is important for both spelling and reading. To determine whether some methods of scoring children's early spellings predict later spelling performance better than do other methods, we analyzed data from 374 U.S. and Australian children who took a 10-word spelling test at the end of kindergarten (mean age 6 years, 2 months) and a standardized spelling test approximately two years later. Surprisingly, scoring methods that took account of phonological plausibility did not outperform methods that were based only on orthographic correctness. The scoring method that is most widely used in research with young children, which allots a certain number of points to each word and which considers both orthographic and phonological plausibility, did not rise to the top as a predictor. Prediction of Grade 2 spelling performance was improved to a small extent by considering children's tendency to reverse letters in kindergarten.

Keywords

spelling; orthographic learning; orthography; phonological processing; longitudinal

Correspondence should be directed to Rebecca Treiman, Campus Box 1125, St. Louis MO 63130. rtreiman@wustl.edu.
Rebecca Treiman and Brett Kessler, Department of Psychological and Brain Sciences, Washington University in St. Louis; Tatiana Cury Pollo, Department of Psychology, Universidade Federal de São João del-Rei; Brian Byrne, School of Behavioural, Cognitive and Social Sciences, University of New England and Australian Research Council Centre of Excellence in Cognition and its Disorders; National Health and Medical Research Council Centre of Excellence in Twin Research; Richard K. Olson, Department of Psychology and Neuroscience, University of Colorado.

Learning to spell is a foundation for both writing and reading (e.g., Graham, Harris, & Chorzempa, 2002; Treiman, 1998), and so it is important to find accurate and sensitive ways to assess children's spelling. The most obvious way to score spellings is as correct or incorrect. However, this type of scoring overlooks the difference between incorrect spellings that reveal some knowledge of sound-spelling correspondences, such as ⟨kam⟩ for *come*, and incorrect spellings that reveal little or no knowledge of these correspondences, such as ⟨bwya⟩ for *come*. Partial spellings such as ⟨kam⟩ and ⟨k⟩ for *come* are common among young children, and research has documented the phonological basis of such spellings (e.g., Read & Treiman, 2013). Measures of spelling performance that take these phonological factors into account may therefore be more informative than is conventional correctness. As Ritchey, Coker, and McCraw (2010) stated in a discussion of kindergartners' spelling,

... an assessment procedure that focuses solely on students' complete spelling abilities (e.g., correct vs. incorrect spelling) may limit the information that can be gained. In contrast, more precise analysis of students' spelling skills that examines incomplete or inaccurate spelling may provide additional information. For example, students' partial spelling (e.g., ⟨ct⟩ for *cat*) or substitution of phonetically similar letters (e.g., ⟨kat⟩ for *cat*) may indicate knowledge of sound-spelling relationships. (p. 78)

The goal of the present study was to explore different measures of kindergartners' spelling and to determine whether some measures are better predictors of later spelling performance than is conventional correctness. In particular, are measures that take account of phonological plausibility better predictors than measures that consider orthographic acceptability only? We addressed these questions using data from an ongoing longitudinal study of literacy development in twins, data that have been used in other work to examine genetic and environmental influences on literacy (e.g., Byrne et al., 2005, 2008; Christopher et al., 2015). We employed eight different metrics to score the spellings of 374 U.S. and Australian children who took a widely used 10-word spelling test at the end of kindergarten and a standardized spelling test at the end of Grade 2, and we asked which measures best predicted Grade 2 performance.

Much theory and research suggests that measures of the phonological quality of young children's spellings would be good indicators of their current knowledge and future prospects. For example, stage and phase theories of literacy development (e.g., Ehri, 2015; Frith, 1985) propose that young children rely on phonological processes to spell and read. When spelling a word, according to these theories, children attempt to segment it into phonemes and assign a letter or letter group to each phoneme. Some phonemes have more than one possible spelling, such as ⟨k⟩, ⟨c⟩, and ⟨ck⟩ for /k/ in English, but these theories state that young children do not store information in memory about which correspondences are appropriate for specific words or contexts within words. What is important for beginning spellers, according to these theories, is the ability to use a plausible spelling for each phoneme. Measures that credit plausible but incorrect spellings of phonemes, such as ⟨k⟩ for the first consonant of *come*, should predict later spelling performance well because they pick up children's knowledge of sound-spelling correspondences. Measures that credit only

letters that are conventional for the specific word, treating ⟨k⟩ as no better than ⟨v⟩ for the first consonant of *come*, should be less good predictors.

One purely phoneme-based scoring method classifies spellings as phonologically plausible or implausible. For example, ⟨kumm⟩ for *come* is phonologically plausible because all of its phonemes are represented with letters or letter groups that are used for those phonemes in some English words. The spelling is not penalized by the fact that the letters are not conventional for this word. Landerl and Wimmer (2008) advocated use of such a measure with children in their first year of formal literacy instruction. In their study of 115 German-speaking children who were followed from Grades 1 to 8, they found a correlation of .47 between phonological plausibility of spellings in Grade 1 (mean age 6 years, 11 months) and conventional correctness in Grade 8. On the basis of this result, they suggested that “good competence in translating phonemic segments into a graphic sequence is indispensable” for later spelling (p. 158).

Phonological plausibility is a binary phoneme-based measure; it overlooks differences among implausible spellings. However, a child who produces spellings such as ⟨km⟩ for *come* and ⟨te⟩ for *tree*, representing some but not all of a word’s phonemes with phonologically appropriate letters, may be more advanced than a child who produces spellings such as ⟨bwya⟩ and ⟨lov⟩ for those words. Correspondingly, the former child may be more advanced when tested several years later. Such considerations have led researchers to develop nonbinary phoneme-based scoring systems that credit plausible spellings of phonemes regardless of whether the letters are conventionally correct. One such system, which Clemens, Oslund, Simmons, and Simmons (2014) called *sound spelling*, allots 1 point to each phoneme that is spelled in a phonologically plausible manner. Caravolas, Hulme, and Snowling (2001) used a more elaborate phoneme-based scoring system, allotting up to 4 points for each phoneme. When the spellings of 148 reception year children in England (mean age 5 years, 1 month) were scored using this system, the correlation with performance on a standardized spelling test in Year 2 was .53. The correlation was somewhat lower, .47, when reception year spellings were scored as correct or incorrect.

In the nonbinary scoring methods discussed so far, phonologically plausible spellings receive credit regardless of whether they are orthographically correct. For example, ⟨f⟩ receives credit for /f/ whether the conventional spelling in a particular word is ⟨f⟩ or ⟨ph⟩. Other nonbinary methods consider both orthographic and phonological acceptability. These methods, which we call *mixed* methods, were popularized by Liberman, Rubin, Duquès, and Carlisle (1985). A certain number of points is typically allotted to each word, and children receive full points for a correct spelling. Children receive fewer points for a spelling that uses phonologically plausible but incorrect letters and still fewer points for a spelling that represents only some phonemes. When McBride-Chang (1998) scored the spellings of 93 U.S. children in the second half of kindergarten using a mixed method, she found a correlation of .50 with performance on a standardized spelling test 10 months later. Although no studies, to our knowledge, have compared mixed scoring to other types of scoring for ability to predict later spelling performance, mixed methods are popular in studies with young children. This is shown by a literature review that we conducted using PsycINFO with the keywords “spelling” and “kindergarten.” We found 50 studies published

in peer-reviewed journals between 2000 and 2015 that classified the spelling of English words by kindergartners in the U.S., Australia, or Canada. We examined the scoring methods used in these studies, finding 28 instances of mixed scoring, 19 of binary correctness, and 15 of other methods (some of the studies included more than one scoring method). Mixed scoring is also popular in studies of preschool children. In a literature review that replaced the “kindergarten” keyword with either “preschool” or “prekindergarten,” we found 11 studies that used mixed methods, 6 that used binary correctness, and 1 that used another method.

A different approach to scoring is to credit only orthographically correct letters. We refer to such scoring systems as *letter based*. We have already discussed correctness, a letter-based binary scoring system. One letter-based nonbinary system, which Clemens et al. (2014) called the *letter sequence* measure, allots 1 point for using the correct first letter of a word, 1 point for the correct last letter, and 1 point for each correct two-letter sequence. Based on the research and theory we have discussed, letter sequence scoring of young children’s spelling should predict later spelling performance less well than do scoring methods that take account of phonological plausibility, either phoneme-based methods or mixed methods. This is because letter sequence scoring does not give credit for orthographically incorrect but phonologically plausible segments. For example, spellings of *come* that begin with ⟨k⟩ are downgraded as much as spellings that begin with ⟨v⟩ or ⟨l⟩, even though ⟨k⟩ is a phonologically plausible spelling of the word’s first phoneme. Clemens et al. reported that spelling performance at the end of kindergarten as assessed by letter sequence scoring showed a positive correlation with reading ability one year later and that the letter sequence measure performed similarly to other nonbinary measures. To our knowledge, researchers have not examined the ability of letter sequence scoring to predict later spelling performance.

We included in the present study three nonbinary scoring measures that are based on the concept of edit distance. This is a way of determining how dissimilar two strings are from one another by counting the minimum number of operations required to transform one string into the other (see Kruskal, 1983). The *phoneme distance* measure counts the number of deletions, additions, and substitutions that are needed to transform a child’s letter string into a phonologically plausible representation of the target. For example, ⟨kmp⟩ requires one addition and one deletion to turn it into ⟨kum⟩, a phonologically plausible spelling of *come*. *Letter distance* counts the number of deletions, additions, and substitutions of single letters that are needed to transform a child’s letter string into the conventional spelling. For example, ⟨sad⟩ for *said* requires one additional letter to make it correct. A penalty value is assigned to each transformation. The number of penalty points reflects the quality of a child’s spelling, with higher values indicating lower-quality spellings. Here we used the same penalties for the phoneme distance and letter distance measures, allowing a good test of whether scoring methods that credit phonologically plausible spellings even when they are not conventional are better predictors of later performance than are purely letter-based methods. Our final measure based on edit distance was AMPR (automated measure of phoneme representation; Treiman & Kessler, 2004). The design of this system took account of common phoneme–letter correspondences and common sound-based errors of young children, such as the tendency to represent the affrication of /t/ before /ɹ/ by producing

spellings such as ⟨chre⟩ for ⟨tree⟩. Several studies of children's spelling have used measures based on edit distance (e.g., Kessler, Pollo, Treiman, & Cardoso-Martins, 2013; Pollo, Kessler, & Treiman, 2009; Zhang & Treiman, 2015), but the ability of these measures to predict later spelling performance has not been examined.

To summarize to this point, surprisingly little research has examined whether some methods of scoring young children's spellings provide more information about later spelling performance than do others. To address this question, we scored kindergartners' spellings on a commonly used 10-word spelling test using the methods that have been used in past studies—correctness, phonological plausibility, correct sounds, mixed, letter sequence, phoneme distance, letter distance, and AMPR—and we examined the relation between these measures and Grade 2 spelling performance. We tested the hypothesis that measures that consider phonology, including those that are based purely on phonemes (phonological plausibility, correct sounds, phoneme distance, AMPR) and those that consider phonological plausibility as well as orthographic correctness (mixed), are better predictors of Grade 2 performance than are purely letter-based measures (correctness, letter sequence, letter distance). We also asked whether nonbinary measures (correct sounds, mixed, letter sequence, phoneme distance, letter distance, AMPR) predict later performance better than do binary measures (correctness, phonological plausibility).

Our main interest was in the letters that children chose to write words, but a secondary interest was in how children wrote these letters. There are many reports of young children producing left-right reversals of letters, for example writing ⟨p⟩ as ⟨q⟩ or writing ⟨j⟩ with the tail opening to the right rather than the left (e.g., Mann, 1993; Mann, Tobin, & Wilson, 1987). A longstanding view is that such errors are a sign of dyslexia (e.g., Terepocki, Kruk, & Willows, 2002). However, several studies have reported that kindergartners' tendency to reverse letters when writing words or printing individual letters has little or no association with later reading skill (Mann, 1993; Mann et al., 1987; Treiman, Gordon, Boada, Peterson, & Pennington, 2014). No study, to our knowledge, has examined reversals in young children's spelling as a predictor of later spelling performance. In the present study, we did not penalize reversals when scoring kindergarten spellings on the correctness, phonological plausibility, sound spelling, mixed, letter sequence, phoneme distance, letter distance, and AMPR metrics. We asked whether supplementing scores on these metrics with a measure of reversals improved the prediction of Grade 2 spelling performance.

Method

Participants

We analyzed data from 158 U.S. and 216 Australian children who had data on both kindergarten and second-grade spelling tests. The sample included both children from 88 same-sex dizygotic twin pairs (38 pairs of girls) and 98 monozygotic twin pairs (48 pairs of girls) and also two girls from a triplet. The U.S. children were recruited from records of twins born in Colorado, and most of the Australian children were recruited from the voluntary Australian Twin Registry. English was the first language of all children. Mean age was 6 years, 2 months at the kindergarten test ($SD = 4.14$ months) and 8 years, 1 month at

the Grade 2 test ($SD = 4.76$ months). In Colorado, at the time that participants were tested, kindergarten was generally a half-day program. In Australia, it was a full day.

Materials and Procedure

Kindergarten—The spelling test was part of a battery of tests given in a session of approximately one hour that took place at the children's homes or schools. For U.S. children, this occurred during the summer following kindergarten. In Australia, testing began during the final two months of the school year. The kindergarten spelling test, which was introduced by Byrne and Fielding-Barnsley (1993) and later used in other studies with children at this level (e.g., Al Otaiba et al., 2010; Kim, Al Otaiba, & Wanzek, 2015), included the words *dog*, *man*, *one*, *said*, *blue*, *come*, *plug*, *went*, *limp*, and *tree*. The words were presented in that order. Each word was said once, used in a sentence, and said again. Children spelled four nonwords after the words, but we do not include these in our analyses because nonwords do not have conventionally correct spellings and so do not permit comparison of letter-based and phoneme-based scoring methods.

Grade 2—The spelling subtest of the Wide Range Achievement Test–Revised (WRAT-R; Jastak & Wilkinson, 1984) was given during an approximately one-hour session that included a number of other tests. The U.S. children were tested in the summer after the end of the second grade. The Australian children were tested during the final three to four months of the school year. The WRAT-R spelling subtest includes 45 words, ranging from easy ones like *go* to difficult ones like *belligerent*. Children spell words until they make ten consecutive errors. Our outcome measure was the number of correct spellings. Letter reversals, which were rare, were counted as incorrect.

Transcription and Scoring of Kindergarten Spellings

The spelling of each word was transcribed, and a reversed or rotated form that was not a real letter was counted as the corresponding letter. Cases that involved left-right reversals, such as a ⟨j⟩ with the tail opening to the right, were included in the reversal count that is described later. When a child used ⟨b⟩, ⟨d⟩, or ⟨q⟩ for a word that contained *d*, *b*, or *p*, respectively, the child was credited with having used the correct letter but as having made a reversal. Those rare forms that were too distant from any letter to be identified were not included in the transcription. A second judge who coded the spellings of 37 randomly chosen children agreed with the primary judge on 98% of the decisions about letter identity for children in each country.

We scored each spelling according to each metric and calculated the total score across the 10 words. The scoring systems are described below and in Table 1, which provides examples of the application of each system. The mixed scoring was done by trained experimenters. The letter distance, phoneme distance, and AMPR scoring was done using the computer program Ponto (Kessler, 2009), and the other scoring systems were computer scored as well. Table 2 shows the minimum and maximum possible scores on each measure.

Correctness—Each item was scored as conventionally correct or not. Items that were not attempted by a child, which constituted about 0.5% of the total, were scored as incorrect, and similarly for the other analyses.

Phonological plausibility—A spelling was scored as phonologically plausible if each phoneme was transcribed using a letter or letter group that may be used to represent the phoneme in General American English, for U.S. children, or Australian English, for Australian children, and if the phonemes were represented in the correct order. We based our scoring on a comprehensive list of the printed words that occur in kindergarten and first-grade reading materials (Zeno, Ivens, Millard, & Duvvuri, 1995), analyzing the words from that list that have a monosyllabic pronunciation in a pronouncing dictionary appropriate to the dialect: Carnegie Mellon Pronouncing Dictionary (Carnegie Mellon University, 1998) for General American English and Unisyn Lexicon (Fitt, 2008) for Australian English. We developed a set of phoneme–letter alignments for the words in each dialect, the minimal set needed to give reasonable, complete alignments. All letters were treated as if they spell some sound, either by themselves or in combination with adjacent letters. For example, the correspondences for /m/ included ⟨m⟩ (as in *mud*), ⟨mb⟩ (as in *comb*), ⟨me⟩ (as in *come*), and ⟨mm⟩ (as in *hummed*). The correspondences for /m/ were the same in American and Australian English, but there were some differences for other phonemes.

Sound spelling—This system (e.g., Clemens et al., 2014; Ritchey et al., 2010) allots 1 point for each phoneme that is spelled using a mapping that is appropriate for the child's dialect. We used the same mappings as for the phonological plausibility measure. As in previous studies using this scoring system, we considered a phoneme to be represented even if its spelling was misordered relative to the spellings of other phonemes. Extraneous letters did not detract from the score.

Mixed—We used the system of Byrne and Fielding-Barnsley (1993) according to which a correct spelling of a word receives 6 points. Five points are given to a spelling that represents all phonemes with plausible letters but where some letter choices are not conventionally correct. A spelling that is judged to represent all phonemes but to include one more distantly related spelling, such as ⟨ch⟩ for the first sound of *tree*, receives 4 points. Productions that represent more than one but not all phonemes receive 3 points. Children receive 2 points if they spell one phoneme with a conventional letter, 1 point if they spell one phoneme with a related letter, and 0 points if they do none of these. A second judge who coded the spellings of 37 randomly chosen children gave the same rating as the primary judge to 88% of the words (87% for U.S. children, 90% for Australian children).

Letter sequence—This system (e.g., Clemens et al., 2014) allots 1 point for use of the correct first letter of the word, 1 point for the correct last letter, and 1 point for each correct two-letter sequence that is produced in the correct order.

Phoneme distance—We determined the phonologically plausible spellings of each word using the correspondences described earlier and measured the distance of the child's spelling from each phonologically plausible spelling. We set a penalty of 1 point for additions, 1 point for deletions, and 1.4 points for substitutions, and we required letters to be in the

correct order. For a child's spelling of each word, we used the best distance score, that is, the lowest.

Letter distance—We determined the number of deletions, additions, and substitutions needed to transform the child's spelling into the conventional spelling. The penalties were the same as for the phoneme distance measure, and correct ordering was required.

AMPR—Using the correspondences from Treiman and Kessler (2004), we determined the number of additions needed to transform the child's spelling of each word into a phonologically plausible spelling. Letters were not required to be in the correct sequence, and extra letters were not penalized.

Reversals—For letters whose standard forms are asymmetrical, we scored productions with left-right orientation errors as reversals. We calculated, for each child, the proportion of reversed letters out of all asymmetrical letters the child produced. The second judge agreed with the primary judge on 99% of the decisions about whether a letter was reversed, with similarly high reliability for the children in each country.

Results

Table 2 shows the mean and standard deviation on each measure for US and Australian participants separately and for all 374 children combined. The Australian children performed significantly better than the US children on the kindergarten measures ($p < .001$, two tailed, for each measure according to mixed-model analyses that used twin pair as a random factor; note that lower scores on the letter distance, phoneme distance, and AMPR measures correspond to better performance). The difference between the two groups on the Grade 2 spelling test did not reach statistical significance ($p = .053$, two tailed).

Table 3 shows the correlations among the measures for all 374 children. The reversal measure was log transformed for this and subsequent analyses because of its large positive skew. All of the kindergarten spelling measures correlated significantly with one another, with the absolute values of the correlation coefficients ranging from .68 to .99. Moreover, all of the kindergarten spelling measures correlated significantly with second-grade spelling performance. Measures that took phonological plausibility into account did not correlate more highly with later spelling performance than did measures based purely on orthographic acceptability. In fact, the phoneme-based measures showed the lowest correlations with second-grade spelling performance (range .48 to .55), the mixed measure was intermediate (.57), and the letter-based measures had the highest correlations (.62 to .63). For the two pairs of letter-based and phoneme-based measures that were most similar, namely correctness and phonological plausibility and letter distance and phoneme distance, the phoneme-based measure correlated significantly less highly with second-grade spelling performance than did the letter-based measure ($p < .001$, two tailed, for both comparisons using the test for the difference between dependent correlation coefficients described by Steiger, 1980). Separate analyses of the data from each country showed the same pattern of lower correlations with second-grade spelling performance for phoneme-based measures than letter-based measures.

To further examine the relations between kindergarten spelling and Grade 2 performance, and to address our secondary question of whether consideration of reversal errors improves the prediction of Grade 2 spelling, we conducted a series of mixed-model analyses. Each analysis included one kindergarten spelling measure, reversals, and country as predictors of Grade 2 performance; twin pair was a random factor. As Table 4 shows, reversals had a statistically significant effect in seven of the eight analyses. The direction of the effect was such that children who reversed more letters in kindergarten tended to be poorer spellers two years later than children who reversed fewer letters. Note that 52% of the children produced at least one reversal in kindergarten; these errors were not limited to a small proportion of children who were destined to become very poor spellers. There was also a tendency, statistically significant in two of the analyses, for US children to perform better on the Grade 2 spelling test than anticipated given their kindergarten performance. This tendency reflects the fact that US children were significantly poorer spellers than Australian children in kindergarten but that there was only a trend in this direction in Grade 2. Table 4 also shows the proportion of variance accounted for by the fixed and random effects in each analyses, using the method of Nakagawa and Schielzeth (2013). The predictive value was lowest for analyses that included purely phoneme-based measures, intermediate for the analysis using the mixed measure, and highest for analyses using letter-based measures. Although models that included reversals and country accounted for significantly more variance in Grade 2 spelling performance than models that did not ($p < .05$ according to likelihood ratio tests), the proportion of additional variance accounted for was never more than 1%.

To probe the limits of the findings, we conducted a series of analyses successively removing the data from children who produced 10 correct spellings in kindergarten, 9 correct spellings, and so on. For each group of children we conducted a series of mixed-model analyses to predict Grade 2 spelling performance from each kindergarten spelling measure, using twin pair as a random factor. Table 5 shows the total proportion of variance explained by the fixed and random factors in these analyses. The pattern noted previously, such that correctness was a better predictor than phonological plausibility and letter distance was a better predictor than phoneme distance, held true for most groups. The only group of children for which the analysis using phoneme distance was numerically superior to the analysis using letter distance was the group with no correct spellings in kindergarten. This group was small, however, and the other nonbinary phoneme-based measures—sound spelling and AMPR—did not perform particularly well for these children. The results in Table 5 further show that binary measures of kindergarten spelling fared increasingly poorly relative to nonbinary measures as the analyses involved poorer and poorer spellers.

As another way to probe the limits of the findings, we examined the results for words with more and less predictable spellings. We classified ⟨blue⟩, ⟨come⟩, ⟨one⟩, ⟨said⟩, and ⟨tree⟩ as having less predictable spellings because they included phonemes with uncommon spellings or long vowels, which may be spelled in multiple ways. ⟨Dog⟩, ⟨limp⟩, ⟨man⟩, ⟨plug⟩, and ⟨went⟩ were classified as having more predictable spellings. As expected, kindergarten scores on letter-based measures were poorer for words in the less predictable category than for words in the more predictable category ($p < .01$ for correctness and letter distance according to mixed-model analyses with twin pair and word as random factors and word type as a fixed factor; $p < .05$ for letter sequence, scoring performance on each word as

a proportion of the maximum possible score on the word). There were no significant differences between the two sets of words on phoneme-based measures. Table 6 shows the correlations among the measures for each word set. Correlations involving phoneme-based and letter-based measures were lower than correlations involving the same type of measure, but this pattern was more pronounced for words with less predictable spellings than for words with more predictable spellings. Even for words with less predictable spellings, however, correlations between letter-based and phoneme-based measures were statistically significant. The pattern noted previously, such that letter-based scoring of kindergarten spelling was most closely related to Grade 2 performance, mixed scoring was intermediate, and phoneme-based scoring was least closely related, was largely carried by words with less predictable spellings. For these words, the letter-based measures of correctness and letter distance correlated significantly more highly with Grade 2 spelling performance than did the otherwise similar phoneme-based measures of phonological plausibility and phoneme distance ($p < .05$, two tailed, for the comparison between correctness and phonological plausibility; $p < .001$, two tailed, for the comparison between letter distance and phoneme distance). These differences were not significant for words with more predictable spellings.

Discussion

Young children often spell words incorrectly. However, use of partial spellings (e.g., ⟨cm⟩ for *come*) and spellings that are phonologically plausible but unconventional (e.g., ⟨foan⟩ for *phone*) suggests that children often have some knowledge about the links between sounds and spellings. Based on the widespread view that phonology is critical for early spelling (e.g., Ehri, 2015; Frith, 1985), one would expect measures of early spelling performance that are based on phonological acceptability to be good indicators of children's current knowledge and future prospects. Measures that penalize orthographically correct spellings even when they are phonologically acceptable should be less good predictors. However, no study has comprehensively tested the idea that measures of early spelling that consider phonological plausibility are better predictors of later spelling performance than are measures based only on orthographic acceptability. We performed such a test, finding some results that were expected and others that were unexpected.

One finding that was expected on the basis of previous research was that spelling performance at the end of kindergarten, regardless of how it was scored, correlated significantly with spelling performance at the end of Grade 2. This result, together with previous findings (Caravolas et al., 2001; Caravolas et al., 2012; Klicpera & Schabmann, 1993; Landerl & Wimmer, 2008; Maughan et al., 2009; McBride-Chang, 1998), shows that there is some stability in the rate of spelling development from an early age.

An unexpected finding of our study, given influential phase and stage theories of literacy development (e.g., Ehri, 2015; Frith, 1985), is the good predictive power of letter-based measures that penalized incorrect choices that were phonologically plausible as much as incorrect choices that were phonologically implausible. Although our spelling test did not include a large number of words, as appropriate given the age of the children, the difference between letter-based and phoneme-based measures seemed to be most apparent for the five words with less predictable spellings, where the two types of measures could be most clearly

differentiated. These results suggest that, by the end of kindergarten in the populations studied here, children have some knowledge about which spellings are conventional for specific words or specific positions in words and that they use this knowledge in spelling production. The results support the emerging view that even young children can attend to and remember visual orthographic features of words (Cassar & Treiman, 1997; Martinet, Valdois, & Fayol, 2004; Treiman & Kessler, 2014; Wright & Ehri, 2007). Children who do this more effectively are more advanced in spelling development than children who do this less effectively and, consequently, tend to perform at an advanced level on later spelling tests. The superiority for letter-based measures over phoneme-based measures was less apparent among the poorer spellers in our study than among the better spellers. Even among children who spelled no words correctly at the end of kindergarten, however, phoneme-based measures did not clearly outperform letter-based measures in the prediction of Grade 2 spelling performance. If there is a point at which spelling is purely based on phonology, it must be quite early in development. Future studies should test this idea by following children from a younger age than done here and comparing nonbinary letter-based and phoneme-based scoring methods.

A second unexpected finding of our study is that, for all children and all words as a whole, the traditional binary measure of correctness fared rather well as a predictor of Grade 2 spelling performance. It is to be expected that, at some point in the development of spelling skill, correctness on an earlier spelling test would predict correctness on a later spelling test and that scoring of earlier spellings for correctness would be a better predictor than other measures. What is surprising is that this was true as early as the end of kindergarten. Landerl and Wimmer (2008) selected a measure of phonological plausibility when scoring the spellings of Grade 1 learners of German, stating that “after only 8 months of formal instruction, it did not seem appropriate to ask children to spell the words orthographically correctly” (p. 153). Our results suggest, however, that conventional correctness can be a good measure of spelling in children who have not had very much formal literacy instruction. It is important to note that correctness fared less well relative to other measures among our less skilled spellers than among the group as a whole. Indeed, only nonbinary measures were significant predictors of Grade 2 spelling among children who spelled no words correctly at the end of kindergarten. Nonbinary measures have value as predictors of later spelling performance, but for a more limited period of time than expected.

A third surprising finding of our study is that mixed scoring of kindergartners’ spellings was not the best predictor of later performance. One might have expected such a result given that, as shown by the literature review presented in the Introduction, the mixed method has been the most popular method of scoring kindergartners’ and preschoolers’ spelling in recent years. Although mixed scoring of kindergarten spelling did predict later spelling performance, consistent with the results of McBride-Chang (1998), it was not the best predictor either with the group as a whole or children who spelled few or no words correctly at the end of kindergarten. The predictive value of mixed scoring might be improved by increasing the bonus that is allotted to fully correct spellings relative to spellings that include phonetically plausible but incorrect letters, but this remains to be investigated. For now, we suggest that researchers consider the letter sequence and letter distance methods, which can be readily computed and do not need to be specially tailored for each language or dialect.

The fact that we found some unexpected results does not seem to be attributable to an unusual sample of children. Other studies that used the same kindergarten spelling test used here and that scored the results using the same mixed method found similar mean scores (Al Otaiba et al., 2010). Our participants were atypical, admittedly, in that they were twins. However, our twin sample was broader than the samples in many other studies in that the children were not all from a single school, city, or country. Importantly, our statistical analyses acknowledged the nesting of children within dyads.

Although our main interest was in the letters that children used to spell words, we were also interested in whether children's tendency to reverse letters in kindergarten accounted for variance in Grade 2 spelling after performance on the kindergarten spelling measures was taken into account. We found, as in previous studies, that children who were poorer spellers at the end of kindergarten were more likely to reverse letters than children who were better spellers (Mann, 1993; Mann et al., 1987). Once performance on the spelling measures was considered, reversal errors made a small but generally statistically significant additional contribution to the prediction of Grade 2 spelling performance. Some previous studies have found little or no relation between reversal errors in spelling words or printing individual letters and later reading ability (Mann, 1993; Mann et al., 1987; Treiman et al., 2014). The present results, with a larger sample of children, suggest that there is some association between reversal errors and later spelling performance but that it is small. In seeking to identify children who are likely to become poor spellers, we can gain more information by considering which letters children choose to spell words than by considering how often they reverse the letters.

Our study included children from US and Australia. Although we found similar patterns of results in the two countries, kindergarten spelling performance was on average better in Australian children. This difference may reflect the fact that the US children in our study generally attended half-day kindergartens whereas the Australian children attended for full days. Also, the Australian families, having voluntarily affiliated with a program of research, may have been more educationally involved than the US families.

A good measure of children's early spelling performance should help forecast future spelling performance, and that was our focus here. A measure should also be sensitive to improvement across time or as a result of intervention. The ability to diagnose an individual child's strengths and weaknesses is another desirable feature. Additional work is needed to explore these characteristics for the spelling measures that we examined. Our exploration of the predictive value of different scoring methods produced some surprising results, ones that do not favor continued reliance on the currently popular mixed method. It will be important to determine how this and other methods fare when tested in other ways and with other groups of children.

Acknowledgments

We thank Kelly Bolland, Kim Corley, and Suzanne Schechtman for their help with the analyses and members of the Reading and Language lab for comments on a draft of the manuscript. This research was supported in part by NIH grants HD051610, HD038526, and HD027802, Australian Research Council grants A79906201, DP0663498, and DP0770805, and NSF grant BCS-1421279.

References

- Al Otaiba S, Puranik CS, Rouby DA, Greulich L, Sidler JF, Lee J. Predicting kindergarteners' end-of-year spelling ability based on their reading, alphabetic, vocabulary, and phonological awareness skills, as well as prior literacy experiences. *Learning Disability Quarterly*. 2010; 33:171–183. DOI: 10.2307/25701446 [PubMed: 25221382]
- Byrne B, Coventry WL, Olson RK, Hulslander J, Wadsworth S, DeFries JC, ... Samuelsson S. A behaviour-genetic analysis of orthographic learning, spelling and decoding. *Journal of Research in Reading*. 2008; 31:8–21. DOI: 10.1111/j.1467-9817.2007.00358.x
- Byrne B, Fielding-Barnsley R. Evaluation of a program to teach phonemic awareness to young children: A 1-year follow-up. *Journal of Educational Psychology*. 1993; 85:104–111. DOI: 10.1037/0022-0663.85.1.104
- Byrne B, Wadsworth S, Corley R, Samuelsson S, Quain P, DeFries JC, ... Olson RK. Longitudinal twin study of early literacy development: Preschool and kindergarten phases. *Scientific Studies of Reading*. 2005; 9:219–235. DOI: 10.1207/s1532799xssr0903_3
- Caravolas M, Hulme C, Snowling MJ. The foundations of spelling ability: Evidence from a 3-year longitudinal study. *Journal of Memory and Language*. 2001; 45:751–774. DOI: 10.1006/jmla.2000.2785
- Caravolas M, Lervåg A, Mousikou P, Efrim C, Litavský M, Onichie-Quintanilla E, ... Hulme C. Common patterns of prediction of literacy development in different alphabetic orthographies. *Psychological Science*. 2012; 23:678–686. DOI: 10.1177/0956797611434536 [PubMed: 22555967]
- Carnegie Mellon University. Carnegie Mellon pronouncing dictionary. 1998. Retrieved May 28, 2003 from <ftp://ftp.cs.cmu.edu/project/speech/dict/cmudict.0.6>
- Cassar M, Treiman R. The beginnings of orthographic knowledge: Children's knowledge of double letters in words. *Journal of Educational Psychology*. 1997; 89:631–644. DOI: 10.1037/0022-0663.89.4.631
- Christopher ME, Hulslander J, Byrne B, Samuelsson S, Keenan JM, Pennington B, ... Olson RK. Genetic and environmental etiologies of the longitudinal relations between pre-reading skills and reading. *Child Development*. 2015; 86:342–361. DOI: 10.1111/cdev.12295 [PubMed: 25263167]
- Clemens NH, Oslund EL, Simmons LE, Simmons D. Assessing spelling in kindergarten: Further comparison of scoring metrics and their relation to reading skills. *Journal of School Psychology*. 2014; 52:49–61. DOI: 10.1016/j.jsp.2013.12.005 [PubMed: 24495494]
- Ehri, LC. How children learn to read words. In: Pollatsek, A.; Treiman, R., editors. *The Oxford handbook of reading*. New York: Oxford University Press; 2015. p. 293-310.
- Fitt, S. Unisyn lexicon [Computer software]. 2008. Available at <http://www.cstr.ed.ac.uk/projects/unisyn/>
- Frith, U. Beneath the surface of developmental dyslexia. In: Patterson, K.; Marshall, JC.; Coltheart, M., editors. *Surface dyslexia: Neuropsychological and cognitive studies of phonological reading*. London, England: Lawrence Erlbaum Associates; 1985. p. 301-330.
- Graham S, Harris KR, Chorzempa BF. Contribution of spelling instruction to the spelling, writing, and reading of poor spellers. *Journal of Educational Psychology*. 2002; 94:669–686. DOI: 10.1037/0022-0663.94.4.669
- Griffin D, Gonzalez R. Correlational analysis of dyad-level data in the exchangeable case. *Psychological Bulletin*. 1995; 118:430–439. <http://doi.org/10.1037/0033-2909.118.3.430>.
- Jastak, SR.; Wilkinson, GS. WRAT-R: Wide range achievement test-revised administration manual. Wilmington, DE: Jastak Associates; 1984.
- Kessler, B. Ponto [Computer software]. 2009. Available at <http://spell.psychology.wustl.edu/ponto>
- Kessler B, Pollo TC, Treiman R, Cardoso-Martins C. Frequency analyses of prephonological spellings as predictors of success in conventional spelling. *Journal of Learning Disabilities*. 2013; 46:252–259. DOI: 10.1177/0022219412449440 [PubMed: 22798104]
- Kim YS, Al Otaiba S, Wanzek J. Kindergarten predictors of third grade writing. *Learning and Individual Differences*. 2015; 37:27–37. <http://doi.org/10.1016/j.lindif.2014.11.009>. [PubMed: 25642118]

- Klicpera C, Schabmann A. Do German-speaking children have a chance to overcome reading and spelling difficulties? A longitudinal survey from the second until the eighth grade. *European Journal of Psychology of Education*. 1993; 8:307–323. DOI: 10.1007/BF03174084
- Kruskal, JB. An overview of sequence comparison. In: Sankoff, D.; Kruskal, JB., editors. *Time warps, string edits, and macromolecules: The theory and practice of sequence comparison*. Reading, MA: Addison-Wesley; 1983. p. 1-44.
- Landerl K, Wimmer H. Development of word reading fluency and spelling in a consistent orthography: An 8-year follow-up. *Journal of Educational Psychology*. 2008; 100:150–161. DOI: 10.1037/0022-0663.100.1.150
- Lieberman, IY.; Rubin, H.; Duquès, S.; Carlisle, J. Linguistic abilities and spelling proficiency in kindergarteners and adult poor spellers. In: Gray, DB.; Kavanagh, JF., editors. *Biobehavioral measures of dyslexia*. Parkton, MD: York Press; 1985. p. 163-176.
- Mann VA. Phoneme awareness and future reading ability. *Journal of Learning Disabilities*. 1993; 26:259–269. DOI: 10.1177/002221949302600406 [PubMed: 8515190]
- Mann VA, Tobin P, Wilson R. Measuring phonological awareness through the invented spellings of kindergarten children. *Merrill-Palmer Quarterly*. 1987; 33:365–391.
- Martinet C, Valdois S, Fayol M. Lexical orthographic knowledge develops from the beginning of literacy acquisition. *Cognition*. 2004; 91:B11–B22. DOI: 10.1016/j.cognition.2003.09.002 [PubMed: 14738774]
- Maughan B, Messer J, Collishaw S, Pickles S, Snowling M, Yule W, MR. Persistence of literacy problems: Spelling in adolescence and at mid-life. *Journal of Child Psychology and Psychiatry*. 2009; 50:893–901. DOI: 10.1111/j.1469-7610.2009.02079.x [PubMed: 19490310]
- McBride-Chang C. The development of invented spelling. *Early Education and Development*. 1998; 9:147–160. DOI: 10.1207/s15566935eed0902_3
- Nakagawa S, Schielzeth H. A general and simple method for obtaining R^2 from generalized linear mixed-effects models. *Methods in Ecology and Evolution*. 2013; 4:133–142. DOI: 10.1111/j.2041-210x.2012.00261.x
- Pollo TC, Kessler B, Treiman R. Statistical patterns in children's early writing. *Journal of Experimental Child Psychology*. 2009; 104:410–426. DOI: 10.1016/j.jecp.2009.07.003 [PubMed: 19691970]
- Read, C.; Treiman, R. Children's invented spelling: What we have learned in forty years. In: Piattelli-Palmarini, M.; Berwick, RC., editors. *Rich languages from poor inputs*. New York: Oxford University Press; 2013. p. 197-211.
- Ritchey KD, Coker DL, McCraw SB. A comparison of metrics for scoring beginning spelling. *Assessment for Effective Intervention*. 2010; 35:78–88. DOI: 10.1177/1534508409336087
- Steiger JH. Tests for comparing elements of a correlation matrix. *Psychological Bulletin*. 1980; 87:245–251. DOI: 10.1037/0033-2909.87.2.245
- Terepocki M, Kruk RS, Willows DM. The incidence and nature of letter orientation errors in reading disability. *Journal of Learning Disabilities*. 2002; 35:214–233. DOI: 10.1177/002221940203500304 [PubMed: 15493319]
- Treiman, R. Why spelling? The benefits of incorporating spelling into beginning reading instruction. In: Metsala, JL.; Ehri, LC., editors. *Word recognition in beginning literacy*. Mahwah, NJ: Lawrence Erlbaum Associates; 1998. p. 289-313.
- Treiman R, Gordon J, Boada R, Peterson RL, Pennington BF. Statistical learning, letter reversals, and reading. *Scientific Studies of Reading*. 2014; 18:383–394. DOI: 10.1080/10888438.2013.873937 [PubMed: 25642131]
- Treiman R, Kessler B. The case of case: Children's knowledge and use of upper- and lowercase letters. *Applied Psycholinguistics*. 2004; 25:413–428. DOI: 10.1017/S0142716404001195
- Treiman, R.; Kessler, B. *How children learn to write words*. New York, NY: Oxford University Press; 2014.
- Wright DM, Ehri LC. Beginners remember orthography when they learn to read words: The case of doubled letters. *Applied Psycholinguistics*. 2007; 28:115–133. DOI: 10.1017/S0142716407070063
- Zeno, SM.; Ivens, SH.; Millard, RT.; Duvvuri, R. *The educator's word frequency guide*. Brewster, NY: Touchstone Applied Science Associates; 1995.

Zhang L, Treiman R. Writing dinosaur large and mosquito small: Prephonological spellers' use of semantic information. *Scientific Studies of Reading*. 2015; 19:434–445. DOI: 10.1080/10888438.2015.1072820 [PubMed: 27064560]

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

Table 1

Descriptions and Examples of Scoring Methods

Method	Type	Description	Examples for <i>tree</i>
Correctness	Binary, letter based	1 for correct spelling of whole word, 0 otherwise	tree:1, trey:0, tee:0, chet:0
Phonological plausibility	Binary, phoneme based	1 if all phonemes represented in correct sequence and no extra letters, 0 otherwise	tree:1, trey:1, tee:0, chet:0
Sound spelling	Nonbinary, phoneme based	1 for each phoneme represented in some position of child's spelling	trey:3, turey:3, tee:2, renoi:2, t:1, h:0
Mixed	Nonbinary, based on both letters and phonemes	6 for fully correct, 5 for phonologically plausible but incorrect spelling, 4 for spelling that represents all phonemes but includes more distantly related letters, 3 for representing more than one but not all phonemes, 2 for representing one phoneme conventionally, 1 for representing one phoneme with a related letter, 0 otherwise	tree:6, tre:5, chree:4, te:3, t:2, ch:1, b:0
Letter sequence	Nonbinary, letter based	1 for correct first letter, 1 for correct last letter, 1 for each correct two-letter sequence	tree:5, tre:4, trea:3, tee:3, try:2, t:1, caby:0
Phoneme distance	Nonbinary, phoneme based	Number of transformations needed to transform child's spelling into a phonologically plausible spelling, counting 1 penalty point for insertions, 1 for deletions, 1.4 for substitutions	tree:0, tre:0, trey:0, triy:1, t:2, tosf:4.8
Letter distance	Nonbinary, letter based	Number of transformations needed to transform child's spelling into correct spelling, counting 1 penalty point for insertions, 1 for deletions, 1.4 for substitutions	tree:0, tre:1, trey:1.4, triy:2.8, t:3, tosf:5.2
AMPR	Nonbinary, phoneme based	Number of additions needed to transform child's spelling into phonologically plausible spelling	chree:0; tie: 2; tosf: 2; m:3

Table 2

Descriptive Statistics for Kindergarten and Grade 2 Measures

Measure	Minimum possible score	Maximum possible score	Mean (SD) scores for participant groups		
			US (n = 158)	Australian (n = 216)	All (n = 374)
Kindergarten					
Correctness	0	10	2.18 (2.01)	3.89 (2.71)	3.17 (2.58)
Phonological plausibility	0	10	3.68 (2.78)	5.49 (2.94)	4.72 (3.01)
Sound spelling	0	33	21.87 (8.60)	26.90 (6.52)	24.78 (7.86)
Mixed	0	60	34.51 (14.34)	45.19 (11.77)	40.68 (13.94)
Letter sequence	0	47	20.66 (10.33)	28.51 (10.14)	25.20 (10.92)
Phoneme distance	0	-- ^a	13.07 (9.27)	8.10 (7.91)	10.20 (8.85)
Letter distance	0	-- ^a	20.72 (9.41)	14.65 (8.84)	17.22 (9.56)
AMPR	0	33	12.35 (8.48)	7.39 (6.33)	9.49 (7.71)
Reversals	0	1	.10 (.14)	.06 (.11)	.07 (.12)
Grade 2					
WRAT correctness	0	45	16.97 (5.55)	18.44 (5.97)	17.82 (5.83)

^aThere is no fixed maximum because a speller can make an unlimited number of additions

Table 3

Correlations among Measures

	1	2	3	4	5	6	7	8	9
1. Kindergarten correctness									
2. Kindergarten phonological plausibility	.88***								
3. Kindergarten sound spelling	.72***	.86***							
4. Kindergarten mixed	.81***	.89***	.97***						
5. Kindergarten letter sequence	.91***	.90***	.92***	.96***					
6. Kindergarten phoneme distance	-.74***	-.89***	-.96***	-.95***	-.91***				
7. Kindergarten letter distance	-.86***	-.88***	-.91***	-.94***	-.97***	.95***			
8. AMPR	-.68***	-.84***	-.99***	-.96***	-.88***	.95***	.88***		
9. Kindergarten reversals	-.32***	-.31***	-.35***	-.35***	-.36***	.31***	.33***	.33***	
10. Grade 2 correctness	.62***	.55***	.52***	.57***	.63***	-.54***	-.62***	-.48***	-.29***

 $p < .001$, one tailed, computed using approach of Griffin and Gonzalez (1995) to take account of nesting of children within pairs

Beta Weights (and Standard Errors) for Kindergarten Spelling Measures, Kindergarten Reversal Errors, and Country in Mixed-model Analyses Predicting Grade 2 Spelling and Proportion of Variance Accounted for by Fixed and Random Factors

Table 4

Kindergarten spelling measure in analysis									
	Correctness	Phono-logical plausibility	Sound spelling	Mixed	Letter sequence	Phoneme distance	Letter distance	AMPR	
Spelling measure	1.41 *** (0.10)	0.98 *** (0.09)	0.36 *** (0.04)	0.24 *** (0.02)	0.35 *** (0.03)	-0.33 *** (0.03)	-0.37 *** (0.03)	-0.33 *** (0.04)	
Reversals	-12.66 * (5.40)	-15.78 ** (5.82)	-14.67 * (5.96)	-13.12 * (5.75)	-10.13 (5.42)	-17.35 ** (5.79)	-14.27 ** (5.43)	-17.62 ** (6.05)	
Country	1.13 (0.60)	0.54 (0.63)	0.57 (0.65)	1.27 * (0.63)	1.41 * (0.59)	0.41 (0.63)	0.99 (0.59)	0.42 (0.66)	
R ²	.66	.59	.59	.61	.66	.59	.65	.56	

* $p < .05$, two tailed,
** $p < .01$, two tailed,
*** $p < .001$, two tailed

Proportion of Variance Accounted for by Fixed and Random Factors in Mixed-Model Analyses Predicting Grade 2 Spelling from Kindergarten Spelling Measure for Groups of Children with Different Numbers of Correct Spellings on Kindergarten Test

Table 5

Group	Kindergarten spelling measure in analysis						
	Correctness	Phono- logical plausibility	Sound spelling	Mixed	Letter sequence	Phoneme distance	Letter distance
All ($n = 374$)	.66***	.58***	.58***	.61***	.66***	.58***	.64***
0–9 correct ($n = 364$)	.63***	.57***	.57***	.60***	.64***	.57***	.62***
0–8 correct ($n = 353$)	.61***	.55***	.57***	.59***	.62***	.57***	.61***
0–7 correct ($n = 340$)	.58***	.53***	.56***	.58***	.60***	.56***	.59***
0–6 correct ($n = 324$)	.55***	.51***	.54***	.56***	.57***	.55***	.57***
0–5 correct ($n = 309$)	.52***	.48***	.52***	.53***	.55***	.52***	.54***
0–4 correct ($n = 283$)	.47***	.45***	.50***	.50***	.52***	.49***	.51***
0–3 correct ($n = 245$)	.48**	.46***	.51***	.52***	.52***	.51***	.52***
0–2 correct ($n = 183$)	.45***	.46***	.49***	.50***	.48***	.49***	.49***
0–1 correct ($n = 106$)	.44**	.40*	.49***	.54***	.49***	.51***	.56***
0 correct ($n = 49$)	--	.01	.08*	.11*	.10*	.16**	.14**

* $p < .05$, two tailed,

** $p < .01$, two tailed,

*** $p < .001$, two tailed

Correlations among Measures for Words with Less Predictable Spellings (Below Diagonal) and Words with More Predictable Spellings (Above Diagonal)

	1	2	3	4	5	6	7	8	9	10
1. Kindergarten correctness		.98**	.83**	.89**	.91**	-.83**	-.85**	-.83**	-.29**	.53**
2. Kindergarten phonological plausibility	.63**		.84**	.89**	.91**	-.85**	-.85**	-.84**	-.28**	.53**
3. Kindergarten sound spelling	.47**	.82**		.97**	.98**	-.98**	-.95**	-.99**	-.33**	.52**
4. Kindergarten mixed	.63**	.83**	.96**		.98**	-.95**	-.94**	-.97**	-.33**	.54**
5. Kindergarten letter sequence	.88**	.75**	.77**	.86**		-.97**	-.97**	-.98**	-.32**	.54**
6. Kindergarten phoneme distance	-.51**	-.88**	-.93**	-.91**	-.75**		.98**	.97**	.30**	-.53**
7. Kindergarten letter distance	-.82**	-.77**	-.78**	-.85**	-.96**	.83**		.95**	.29**	-.53**
8. AMPR	-.31**	-.76**	-.96**	-.90**	-.62**	.89**	.64**		.33**	-.52**
9. Kindergarten reversals	-.28**	-.29**	-.35**	-.35**	-.35**	.29**	.32**	.30**		-.29**
10. Grade 2 correctness	.57**	.49**	.51**	.56**	.63**	-.52**	-.62**	-.41**	-.29**	

** $p < .001$, one tailed, computed using approach of Griffin and Gonzalez (1995) to take account of nesting of children within pairs