2020

Measurement of spelling ability: construction and validation of a phonological, orthographic and morphological pseudo-word instrument for students in Grades 3-6

Tessa Daffern
University of Wollongong, tdaffern@uow.edu.au

Ajay Ramful
Mauritius Institute of Education

Measurement of spelling ability: construction and validation of a phonological, orthographic and morphological pseudo-word instrument for students in Grades 3-6

Abstract

Building on current theoretical understandings of how children learn to spell, this paper reports the design and validation of a new pseudo-word dictation test (labelled the Components of Spelling Test: Pseudo-word version) to measure three spelling components underpinning Standard English: phonology, orthography and morphology. For the first phase of the study, the instrument was tested on a calibration sample of 381 students from Grades 3 to 6, aged between 8 and 12 years. Two versions of the test were recursively developed for Grades 3 and 4 (Pseudo-word-G-3-4) and Grades 5 and 6 (Pseudo-word-G-5-6). In the second phase of the study, the calibrated instrument was validated on a different sample of students in Grades 3 and 4 (n = 224) and Grades 5 and 6 (n = 233). The instrument shows high reliability (0.79-0.92) across the spelling components. A key feature of the instrument is that it affords three specific measures of spelling to align with Triple Word Form Theory. This instrument can be used by teachers to screen students with difficulties in spelling and resultantly plan for targeted instruction in school contexts. It can also be used as a measure of spelling ability for experimental, developmental and correlational research purposes. This novel instrument fills a gap in spelling ability research literature by providing the first pseudo-word metric to assess 8- to 12-year-old students' phonological, orthographic and morphological spelling skills.

Disciplines

Education | Social and Behavioral Sciences

Publication Details

Daffern, T. & Ramful, A. (2020). Measurement of spelling ability: construction and validation of a phonological, orthographic and morphological pseudo-word instrument for students in Grades 3-6. Reading and Writing, 33 571-603.

Measurement of spelling ability: construction and validation of a phonological, orthographic and morphological pseudo-word instrument for students in Grades 3–6

Tessa Daffern[1] · Ajay Ramful[2]

Abstract

Building on current theoretical understandings of how children learn to spell, this paper reports the design and validation of a new pseudo-word dictation test (labelled the *Components of Spelling Test: Pseudo-word version*) to measure three spelling components underpinning Standard English: phonology, orthography and morphology. For the first phase of the study, the instrument was tested on a calibration sample of 381 students from Grades 3 to 6, aged between 8 and 12 years. Two versions of the test were recursively developed for Grades 3 and 4 (*Pseudo-word-G-3-4*) and Grades 5 and 6 (*Pseudo-word-G-5-6*). In the second phase of the study, the calibrated instrument was validated on a different sample of students in Grades 3 and 4 ($n=224$) and Grades 5 and 6 ($n=233$). The instrument shows high reliability (0.79–0.92) across the spelling components. A key feature of the instrument is that it affords three specific measures of spelling to align with Triple Word Form Theory. This instrument can be used by teachers to screen students with difficulties in spelling and resultantly plan for targeted instruction in school contexts. It can also be used as a measure of spelling ability for experimental, developmental and correlational research purposes. This novel instrument fills a gap in spelling ability research literature by providing the first pseudo-word metric to assess 8- to 12-year-old students' phonological, orthographic and morphological spelling skills.

Keywords Spelling ability · Spelling assessment · Phonology · Orthography · Morphology · Linguistics

A condensed version of the preliminary findings of this paper was presented at the European Conference on Educational Research (September, 2018).

✉ Tessa Daffern
tdaffern@uow.edu.au

[1] Faculty of Education, University of Wollongong, Wollongong, Australia

[2] Mathematics Education Department, Mauritius Institute of Education, Moka, Mauritius

The complexity of standard English spelling

Spelling ability is a critical literacy skill of sustained concern among educators, parents and employers as it can support learning to read (Ehri, 2000; Martin-Chang, Ouellette, & Madden, 2014; Moats, 2005/06) and it can impact one's capacity to write (Daffern, Mackenzie, & Hemmings, 2017a; Sumner, Connelly, & Barnett, 2016). Being able to spell in the English language is also a complex linguistic problem-solving process involving integration of phonology, orthography and morphology (Daffern, 2015, 2018; Garcia, Abbott, & Berninger, 2010). For example, to spell an unfamiliar word, phonological processes are activated, requiring awareness of spoken sounds at the smallest speech sound (phoneme) level, as well as at the syllable level. Simultaneously, orthographic processing may be activated and this requires sensitivity to conventional letter strings or patterns within words, including knowing plausible alternative grapheme (alphabetic letter) combinations that apply under positional constraints (Kohnen, Nickels, & Castles, 2009; Treiman & Kessler, 2006). Further, morphological processing may be activated, and this demands sensitivity to the smallest meaningful units in words, such as knowing how suffixes and prefixes attach to base words (Apel, 2014).

Using spelling assessment data to inform teaching priorities

Considering English spelling is underpinned by phonological, orthographic and morphological components, an informative assessment instrument designed to measure English spelling ability should be one that can yield insight into how a student applies such components when spelling. Understanding the types of errors a student makes may assist with the identification of the most effective intervention approach for individual students (Breaux, Bray, Root, & Kaufman, 2017). However, teachers and clinicians such as speech and language practitioners do not have sufficient access to spelling assessment instruments that enable them to provide the explicit and targeted instruction that is needed to improve spelling outcomes (Graham, Harris, & Adkins, 2018; Graham & Santangelo, 2014; Kohnen et al., 2009). Spelling measures are generally presented as a series of dictation tasks whereby children are required to spell words that are orally dictated to them (Breaux et al., 2017), or as tasks that require children to identify and edit spelling errors (see, for example, Australian Curriculum, Assessment, & Reporting Authority (ACARA), 2016). A common scoring method is then typically determined on the accuracy of whole words that are spelled by a student. Yet, the instructional value of a spelling ability instrument is in its capacity to precisely determine which underlying linguistic processes may be impeding spelling accuracy and thus demand instructional priority. A dictation task which provides a framework for "spelling error analysis" can be beneficial "for screening, progress monitoring, and diagnostic purposes" (Al Otaiba & Hosp, 2010, p. 4). Such form of assessment (see, for example, *Words Their Way Inventories*: Bear, Invernizzi, Templeton, & Johnston, 2012; *Single Word Spelling Test*: Sacre & Masterson, 2000) is becoming recognised and valued in school contexts as it can help to identify specific strengths and weaknesses in children's

spelling. By analysing spelling errors, it is possible to understand which cognitive strategies may be applied by a student, and this may provide valuable information about a student's phonological, orthographic and morphological skills (Varnhagen, McCallum, & Burstow, 1997).

Limitations in existing measures of spelling ability

Current instruments of spelling ability that involve error analysis methods have generally offered useful means to understand spelling ability; however, they are limited in varying ways. In reviewing commercially available spelling assessments, Kohnen et al. (2009) concluded that teachers may need to administer several tests in order to determine which spelling components need instructional attention as there is no single instrument that adequately captures all linguistic processes involved in spelling.

Dictation-based instruments which include an analysis of spelling errors can offer a more robust measure of spelling ability than error analysis of words produced in freely composed writing; however, there are limitations inherent in dictation-based and error-analysis instruments which solely rely on students' ability to spell real words. Real-word measures can be problematic because a child might have developed knowledge of the particular word that they have been asked to spell and there is no way of confirming whether the child could spell that word correctly but not yet know the underlying linguistic generalisation that is being assessed (Kohnen et al., 2009). A longitudinal study (Garcia et al., 2010) has shown that spelling pseudowords tends to correlate more than real words with phonological, orthographic and morphological scales when age variations are considered. Indeed, testing spelling using real words may to some extent reflect word-based knowledge rather than knowledge of the linguistic components that underpin spelling.

Another limitation appearing in commonly used spelling assessments which utilize error analysis concerns theoretical alignment (Daffern, 2018). Several existing measures of spelling ability are based on stage theory which implies that phonology, orthography and morphology develop in succession. Currently, there is an increasing realisation that spelling skills develop concurrently along the three dimensions (Bahr, 2015; Daffern, Mackenzie, & Hemmings, 2015; Devonshire, Morris, & Fluck, 2013; Treiman, 2017b), described in the next section. Considering students are capable of integrating phonological, orthographic and morphological skills to spell from the early years of learning to write (Bahr, 2015; Daffern, 2017; Devonshire & Fluck, 2010; Garcia et al., 2010; Rittle-Johnson & Siegler, 1999; Treiman, 2017a; Varnhagen et al., 1997), assessment instruments of spelling ability should include distinct measures of these core linguistic features. Nevertheless, Piagetian notions of spelling development (Gentry, 2000) are reflected in numerous existing measures (Bear et al., 2012; Ganske, 1999), as evidenced in the way spelling ability is classified into a particular developmental stage (Bear & Templeton, 1998) or phase (Ehri, 2005), rather than in terms of the ability to accurately apply phonology, orthography and morphology when spelling. While there is still a need to further our current understandings of how to assess spelling ability (Treiman, 2017a),

existing measures are not sensitive enough to capture specific phonological, orthographic and morphological complexities that are needed to inform instructional priorities (Kohnen et al., 2009).

Triple Word Form Theory

Centred on the notion of phonology, orthography and morphology, Triple Word Form Theory (Bahr, 2015; Daffern, 2018; Daffern et al., 2015; Garcia et al., 2010; Richards et al., 2006) provides a conceptual framework to understand the development of spelling skills. Triple Word Form Theory predicts that the trajectory of learning to spell "depends on learning to code into memory, analyse, and coordinate" phonology, orthography and morphology, and that "children must learn how to cross-map the interrelationships" among these three word forms (Bahr, Silliman, Berninger, & Dow, 2012, p. 1588). As an illustrative example, Fig. 1 shows how phonology, orthography and morphology may be integrated (or cross-mapped) to achieve accurate English spelling. The arrows indicate possible word-form connections that could be made when determining how to spell the word, 'kicked', assuming a child has not yet committed the spelling of this word into long term memory. In this example, in order to correctly spell <kicked>, the child may need to consider the meaning of the word in context and that <-ed> is needed to mark past tense for the verb. Without considering the morphological constituents in the word, <kicked> could be misspelled as <kickt>. In addition, the child may need to mentally segment individual phonemes (e.g., /k/i/k/t = four speech sounds) and consider whether to use the letters <k>, <c> or <ck> (e.g., <**c**icked> and <ki**k**ed> are not orthographically plausible letter patterns in this context). If a breakdown in the

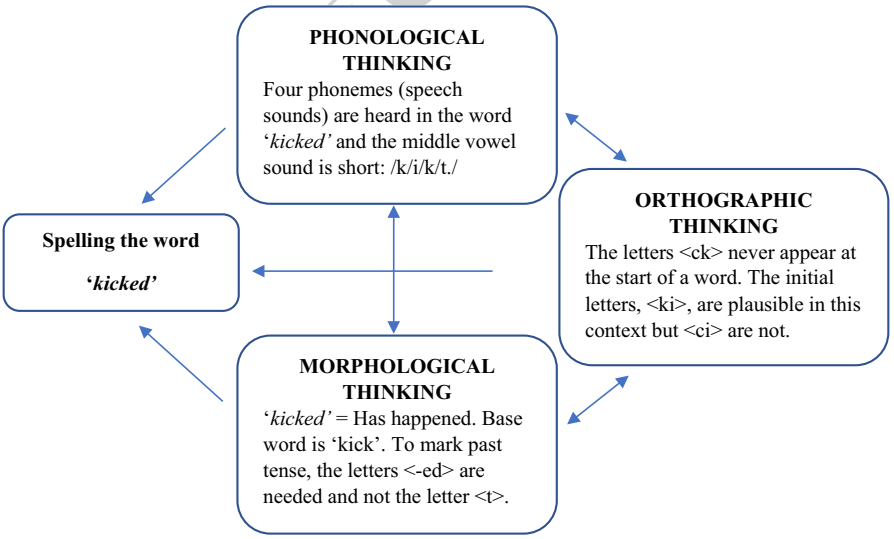

Fig. 1 Relationship among spelling components when spelling the word 'kicked'

coordination of these linguistic processes occurs, the child may misspell the word. If the child is explicitly taught how to coordinate phonology, orthography and morphology, coordination of these word forms may become increasingly efficient and accurate over time.

Phonology

Phonological knowledge encompasses several subskills concerned with the structure of sounds in the spoken language (McLeod & McCormack, 2015). These subskills include awareness of phonemes and how they correspond to graphemes; ability to segment and blend phonemes (for example, /s/t/r/o/ng = /strong/); ability to manipulate onset and rime patterns (for example, /b/-at/;/c/-at/;/m/-at/); ability to substitute or omit phonemes (for example, /sat/without the initial phoneme is/at/); and ability to identify syllables in words. Working memory may influence phonological processes (Daffern, 2017; McLeod & McCormack, 2015). For example, cognitive demand may be high when spelling a word with many syllables because there are a large number of phonological constituents that need to be held in working memory, analysed and then sequentially encoded (Berninger et al., 1998; Larkin, Williams, & Blaggan, 2013; Ruberto, Daigle, & Ammar, 2016). As Larkin et al. (2013) suggest, improvements in the spelling of polysyllabic words, as children mature, could be due to changes that occur in working memory as children get older. Indeed, Gathercole (2007) posits that as efficiency in working memory increases with age, performance in such tasks may improve. Currently available pseudo-word instruments measuring spelling skills may offer useful starting points to determine knowledge of phoneme-grapheme conversion rules (Siegel, 2008); however, they do not distinguish differences in ability to spell phonologically regular words of varying syllable numbers. For example, Kohnen, Colenbrander, Krajenbrink, and Nickels (2015) developed a pseudo-word measure to assess applications of sound-to-letter correspondences; however, the items in their measure are limited to single-syllable pseudo-words. In designing a new measure of spelling ability, items should be sensitive in capturing the phonological complexities in words, at the phoneme and varying syllable levels.

Orthography

Orthographic aspects of spelling are concerned with sub-lexical conventions that are specific to a particular written language (Bowers & Bowers, 2017). Sometimes also referred to as *graphotactic features* (Treiman, 2017a), sub-lexical conventions pertain to the typical arrangement of letter groups (or strings of letters) within words that are present in a writing system (Daffern, 2017). For example, in the standard English writing system, a long vowel phoneme can be represented in several ways (e.g., as in <l**a**t**e**>, <w**ai**t>, <str**aigh**t>, <fr**eigh**t>) but the spelling of some letter patterns can be constrained because they are context sensitive (e.g., the letter sequence, <ou>, rarely appears in the final position of a word where its corresponding phoneme is pronounced/ow/as in the word ‘cow’ (e.g., <gr**ou**nd>, <f**ou**nd>).

Treiman (2017a) proposes that children can pick up information about orthographic patterns through exposure to print; however, explicit instruction is likely to increase the rate of learning.

Morphology

English words are also made up of meaningful units (morphemes). Knowledge of morphemes is important for spelling in English because a phoneme or phoneme sequence may be spelled one way when it is a morpheme (e.g., <ed> for the past tense morpheme, even when this morpheme is pronounced as/t/) yet a different way when it is not a morpheme (e.g., <t> for/t/when/t/does not function as a past tense morpheme). Therefore, accurate spelling requires morphological awareness (Apel, 2014), which is characterized by "sensitivity to the internal, meaning-related structure of words" (Green et al., 2003, p. 752). For example, morphological awareness includes knowing "the meaning of affixes and the alterations in meaning and grammatical class they bring to base words/roots" (Apel, 2014, p. 200) (e.g., knowing that the inflected suffix, <-ed>, in a verb indicates an action in the past tense, such as in <stopped>). Morphological awareness also entails knowing "the manner in which written affixes connect to base words/roots, including changes to those base words/roots" (Apel, 2014, p. 200) (e.g., knowing when a consonant grapheme is doubled such as in <run> to <running>).

Fuelling the research efforts to understand the spelling ability of primary school children on the basis of Triple Word Form Theory, an instrument was recently developed (the *Components of Spelling Test (CoST)*: Daffern et al., 2015). The instrument has been tested and validated with a range of students. The instrument has an internal reliability (Cronbach's alpha) ranging from 0.78 to 0.94. However, the CoST is based on real words only. Kohnen et al. (2009) assert that a real word measure may over-estimate a respondent's spelling ability due to potential influences of prior word-specific knowledge. Thus, the study presented here sought to build a pseudo-word spelling instrument based on the design of the real-word *CoST*. The development of a pseudo-word instrument was also motivated by the need to provide classroom teachers, tutors and specialist clinicians, such as education psychologists or speech and language pathologists, with a supplementary measure of spelling ability that informs teaching priorities and intervention plans, as well as one that can be used to track student learning over time or to determine teaching effectiveness. Importantly, it is not advisable to re-administer the same set of prescribed words from a test (e.g., the real-word version of the *CoST*) multiple times because the students may become familiar with the words being tested. This particularly applies if the purpose of the test is to assess rule-based rather than word-specific knowledge. Test administrators should be mindful that over-use of the same instrument may lead to invalid test results. Adding a pseudo-word instrument to a repertoire of existing forms of spelling assessment can equip teachers or clinicians to better monitor student progress, review teaching approaches, and respond to individual learning needs. Furthermore, as Snowling and Hulme (2012) assert, students displaying difficulty with decoding (reading) also experience spelling difficulties. Thus, the new

test could be used to inform intervention plans for students with persistent spelling and decoding difficulties, including those students who are diagnosed with dyslexia or dysgraphia. Finally, the items considered for inclusion in the new pseudo-word instrument have been designed so that the test can be used in any English-speaking context (e.g., Canada, USA, Great Britain, Australia or New Zealand). This new instrument is available at www.tessadaffern.com.

The study: construction and validation of a pseudo-word instrument

This study presents the design and validation of a new dictation test, labelled the *Components of Spelling Test (CoST): Pseudo-word version*. The design of a new instrument was motivated by the need for a pseudo-word measure that identifies phonological, orthographic and morphological errors in English spelling conventions. The design of this pseudo-word instrument aims to align with the notion that phonological, orthographic and morphological skills are required for Standard English spelling. Thus, we sought to explore if Triple Word Form Theory can be applied in the design of a new pseudo-word measure of spelling ability. In order to test the psychometric properties of the newly designed instrument, the following research questions guided the study:

1. How does spelling ability differ from Grade 3 to Grade 6 (aged 8–12 years) as measured by the phonological, orthographic and morphological scales of the pseudoword *CoST*?
2. How does the newly-designed instrument fare in terms of reliability and validity across Grades 3–6?

Methods

Instrument development

To develop the pseudo-words for the new instrument, the chief researcher (first author) adapted some of the real words from the existing real-word *CoST* by substituting one or more graphemes. This process was supplemented by analyzing the linguistic properties of items within the words of other existing tests and adapting some of those words where appropriate. As in the real-word *CoST*, the pseudo-word version was constructed around three scales, namely the Phonological Component, Orthographic Component and Morphological Component (see Table 1). The pseudo-word instrument provides additional insights about spelling ability that the real-word instrument does not offer. Therefore, administering both the pseudo-word test and the real-word test can be particularly helpful in understanding the nature of a student's difficulty with spelling and then for planning a suitable intervention. To illustrate some of the key differences between the two instruments, Table 1 provides

Table 1 Comparison between real-word and pseudo-word CoST features

Spelling components	Pseudo-word (number of items)	Real-word (number of items)
Phonological	Monosyllabic (10)[a]	Initial & final consonant (5) (e.g. ***t**a**g***)
	Epenthesis	Short vowel graph (5) (e.g. *t**a**g*)
	Elision	Consonant digraph (5) (e.g. ***ch**ew*)
	Substitution	
	Disyllabic (10)[a]	
	Epenthesis	
	Elision	
	Substitution	
	Polysyllabic (10)[a]	Medial blend (16) (e.g. *dia**gnos-tic**i**an*)
	Epenthesis	
	Elision	
	Substitution	
Orthographic	Part A: Constrained letter pattern[b] (28)	Common long vowel (7) (e.g. *sp**ea**ker*)
	Part B: Common long vowel patterns[c] (11)	Ambiguous vowel (7) (e.g. *b**oi**l*)
		Complex consonant patterns (5) (e.g. smu***dge***)
		Syllable juncture consonants (5) (e.g. *bo**tt**le*)
		Unaccented final syllables (5) (e.g. *bottl**e***
Morphological	Inflected suffix[d] (23)	Inflected suffix (7) (e.g. march**ed**)
	Derivational suffix[d] (16)	Derivational suffix (8) (e.g. *opposit**ion**)*
	Prefix[d] [includes non-assimilated & assimilated prefixes (8)	Assimilated prefix (7) (e.g. ***cor-r**espond*)
	Greek and Latin root[e] (8)	Greek and Latin root (7) (e.g. ***chlor**ine*)
		Morpheme juncture schwa vowel (5) (e.g. *opp**o**sition)*
		Homophone (7) (e.g. *w**aist***)

[a]All phonemes are analysed (initial consonants, final consonants, short vowel graphs, consonant blends and consonant digraphs

[b]Only one plausible spelling for each item due to positional constraints

[c]Multiple spelling possibilities are acceptable for each item

[d]Sentences are dictated and all words are visible to the student except for the affixed pseudo-word by which the student is required to spell

[e]The pseudo-word is not visible to the student but the remaining words in the sentence are

Table 2 Phonological features unique to the pseudo-word (PW) instrument

Constructs	Monosyllabic words	Disyllabic words	Polysyllabic words
Encoding short vowel graphemes	N/A	<a>, <e>, <i>, <u>	N/A
Encoding consonant graphemes	<d>, <z>, <s>, <h>, <p>	<l>, <p>, <m>, <b>, <t>, <r>, <y>, <z>, <h>	<m>, <y>
Encoding consonant digraphs	<sh>, <ng>	<sh>, <ng>	<sh>, <ng>
Encoding consonant blends	<sp>, <pl>, <mp>, <nt>, <nd>	<dr>, <bl>, <st>, <nt>, <fl>, <sp>, <mp>, <nk>, <nd>	<gl>, <tr>, <mp>, <dr>, <nd>, <fl>, <pr>

<> indicates alphabetic letters

Table 3 Unique linguistic properties in the orthographic component of the pseudo-word version

Examples of unique orthographic properties	Sample pseudo-words (items underlined and bold)
<k> (when/ki/is heard in initial position, as in 'kiss') <ci> is not plausible	**k**ish (not **c**ish or **ck**ish)
<k> (when/nk/is heard in final position, as in 'sunk') <nc> is not plausible.	blun**k** (not blun**c** or blun**ck**)
<ck> (when/ick/is heard in final position, as in 'stick') <c> or <k> are not plausible	smi**ck** (not smi**c** or smi**k**)
<-ve> (when/v/is heard in final position, as in 'glove'). It is not plausible to end a word in <v>	slo**ve** (not slu**v**)
<dd> (syllable juncture doublets, as in 'puddle'). The consonant doublet is needed because of the short/u/vowel in the first, accented syllable	plu**dd**le (not plu**d**le)
<b> (syllable juncture consonant, as in 'noble'). The medial consonant is not doubled if the vowel in the first syllable is long)	flo**b**le (not flo**bb**le)
<ou> (when the diphthong occurs before/nt/, as in 'mount', or/t/, as in 'shout'). <ow> is not plausible because the phoneme in this word is followed by/nt/	**blou**nt (not blo**w**n**t**)
<ow> (when the diphthong occurs before/n/, as in 'clown', or when the diphthong ends the word, as in 'now'). <ou> is not plausible in these contexts	gl**ow**n (not gl**ou**n)
<ow> or <ough> (when the diphthong ends the word, as in 'now' and 'plough'). <ou> is not typical in this context	sp**ow** or sp**ough** (not sp**ou**)
<oy> (when the diphthong occurs at the end of the word, as in 'toy'). <oi> is not plausible in the final position of a word	z**oy** (not z**oi** or z**oye**)
<oo> or <oul> (when the medial vowel is followed by/d/, as in 'wood' or 'would')	th**oo**d or th**oul**d (not th**o**d**e**)

<> indicates alphabetic letters

//indicates phonemes

an overview of the distinct constructs for both instruments while Tables 2 and 3 include finer details about the unique items in the pseudo-word instrument.

In delineating potential items for the Phonological Component of the new instrument, current literature on phonological processing and on existing

phonologically-based instruments were considered (Daffern et al., 2015; Kohnen et al., 2015; Wagner, Torgesen, Rashotte, & Pearson, 2013). A novel feature of the phonological scale is its capacity to identify spelling accuracy in phonologically regular monosyllabic, disyllabic and polysyllabic pseudo-words. The initial design resulted in the formation of 30 items for the Phonological Component. This component of the pseudo-word instrument builds on the real-word version by providing information about a student's ability to spell regular one-syllable words, two-syllable words and three-syllable words. The real-word instrument does not provide information differentiated by number of syllables. The monosyllabic, disyllabic and polysyllabic dimensions are important for classroom educators and clinical specialists, such as education psychologists or speech and language therapists as they need to determine an appropriate sequence for teaching phonological skills in spelling. For example, if errors are displayed in items within the monosyllabic construct, a focus on learning to spell regular monosyllabic words would be appropriate, before proceeding to disyllabic words and polysyllabic words. Moreover, several phoneme-grapheme correspondences are included in the pseudo-word instrument that are not included in the real-word instrument (see Table 2), thus permitting a more comprehensive insight into a student's phonological applications in spelling. Further, in using the pseudo-word test, qualitative insights can be obtained by determining if errors involve a phonological epenthesis, omission or substitution (Masso & Baker, 2015). However, there may be circumstances where it is not possible to confirm the phonological nature of a substitution or an omission (e.g., if a child spells <tid> as <ted>, the test does not confirm whether the incorrect medial vowel phoneme-grapheme correspondence is due to a difficulty in differentiating the vowel phoneme or a difficulty in applying the correct phoneme-grapheme mapping).

To construct items for the Orthographic Component, existing spelling ability measures were analyzed in order to identify conventional letter patterns for potential inclusion (Conrad, Harris, & Williams, 2013; Daffern, Mackenzie, & Hemmings, 2017b; Kohnen et al., 2015; Kohnen, Nickels, Castles, Friedmann, & McArthur, 2012; Treiman & Kessler, 2006). In developing each pseudo-word for this component, it was ensured that a plausible letter pattern was included in each item. Two sets of letter arrangement patterns involving pseudo-words were constructed for this scale: Part A attempts to capture respondents' awareness of lexical conventions in constrained letter patterns (for example, knowledge of when to use <ou> as opposed to <ow>). Part B assesses knowledge of common long vowel patterns (i.e., plausible letter pattern possibilities for corresponding long vowel phonemes). For each pseudo-word in this scale, only the spelling of specific target letters is assessed rather than the spelling of the whole pseudo-word. The two parts resulted in the formation of 39 items for the Orthographic Component.

In designing the Morphological Component, existing measures which include morphological features (e.g., inflected suffixes) were analyzed in terms of their subconstructs (Bryant & Nunes, 2009; Daffern et al., 2015; Nunes & Bryant, 2006; Nunes, Bryant, & Olsson, 2003). Existing pseudo-word instruments are limited as they only include items featuring inflected suffixes marking tense and plurality and involve speaking and reading tasks rather than spelling tasks. An innovation of the present test design is the inclusion of four morphological constructs using

pseudo-words in a semantic context. In developing the items, a broad range of word types was considered in terms of their phonological form and grammatical function. For each pseudo-word, a sentence was designed using real words but with the pseudo-word embedded in a way that made it functional in context. Each sentence is presented as a cloze activity. The student taking the test writes the target pseudo-word above the line indicated on their response sheet, which includes the sentence that they can read and listen to as they spell the target pseudo-word. The test administrator is required to read the entire sentence, including the pseudo-word that the test-taker is required to spell. For each pseudo-word, only the spelling of specific target letters is assessed rather than the spelling of the whole pseudo-word. The initial design resulted in the formation of 55 items for the Morphological Component.

The Orthographic and Morphological Components of the pseudo-word instrument include a more comprehensive range of linguistic properties than the real-word instrument (Daffern et al., 2015). Tables 3 and 4 include examples of the additional linguistic properties that are unique to the pseudo-word instrument (that is, they do

Table 4 Unique linguistic properties in the morphological component of the pseudo-word version

Examples of unique morphological properties	Sample dictations (items underlined and bold)
Plural suffix from base word ending in <y> (e.g. baby/babies).	Here is one slaby. Here are two slab**ies**
Plural suffix from base word ending in <x> (e.g. box/boxes)	I have one hox. She has ten hox**es**
Plural suffix from consonant–vowel-consonant base word (e.g. bed/beds)	There was one ved. Now there are four ved**s**
Plural suffix from base word containing split/o/ digraph (e.g. drone/drones)	There was one rone. Now there are five ron**es**
Present progressive tense from base word containing split/i/digraph (e.g. slide/sliding)	I will vipe. You are vi**ping**
Past tense from consonant–vowel-consonant base word (e.g. stab/stabbed)	I will clom today. Yesterday, I clo**mmed**
Past participle from base word containing vowel digraph followed by a single consonant (e.g. eat/ eaten)	I will fleat to the shops. He has fleat**en** to the shops
Superlative from base word containing split digraph (e.g. late/latest)	This chair is vate. It is the va**test** chair
Comparative adjective from base word containing short medial vowel followed by final consonant (e.g. big/bigger)	His ball is greb. My ball is gre**bber**
Verb to noun ending in <-er> (e.g. beg/begger)	The man snegs. He is called a sne**gger**
Abstract noun to person noun (e.g. magic/magician)	A person who makes plagic is called a plagi**cian**
Adjective to adverb (e.g. happy/happily)	The dog was greppy. The dog barked grepp**ily**
Non-assimilated prefixes (e.g. mis-; un-; dis; re-).	A person who is not bleam is **un**bleam
Greek and Latin roots (aqua; phobia; sphere; psych; hydro; audio; chrono)	The **aqua**bost ran out of water

<> indicates alphabetic letters

//indicates phonemes

not exist in the orthographic and morphological scales of the real-word version of the *CoST*):

The integration of the different sub-constructs in the initial design of the *CoST: Pseudo-word* instrument resulted in 124 items (see Table 1). The next step was to measure the phonological, orthographic and morphological skills in spelling at each grade level and to identify how the newly developed test items fare in terms of reliability.

Psychometric testing

To test the psychometric properties of the new instrument, data collection and analyses were conducted in two phases. For the first phase, an expert review process occurred (see section on Content Validity) followed by school-based testing using a calibration sample of students in Grades 3–6 (referred to as sample one). Students in sample one (calibration sample) were invited to complete the newly designed instrument and analyses were conducted using their data. This process resulted in item reduction and the development of two versions of the instrument, referred to as *Pseudo-word-G-3-4* (for Grades 3 and 4) and *Pseudo-word-G-5-6* (for Grades 5 and 6) due to differences in students' ability to spell the pseudo-words. For the second phase of the study, data from a different sample of students (referred to as sample two) were used to validate both the *Pseudo-word-G-3-4* and *Pseudo-word-G-5-6* versions.

Participating students

Given that, across grades, students are at different levels of maturity in terms of spelling ability, it was important to test which items can be attempted by the different age groups. Also, as students proceed towards the middle school years, the vocabulary demands increase across subject areas and this can pose challenges with spelling. Hence, an assessment that measures a comprehensive range of linguistic skills in spelling is needed so that teachers and clinicians can help students to build linguistic skills that are essential for writing in Grades 3–6 and beyond. Furthermore, given that the real-word *CoST* had previously been designed and tested for students in Grades 3–6, it was logical to focus the sampling for the present study on the same age group for comparison and correlation purposes.

Sample one

Five schools from a metropolitan city in Australia were involved in calibration testing (referred to as phase one). The schools were chosen from government and Catholic jurisdictions through a convenience sampling method and represented a socio-economic demographic that was marginally higher than the national mean for Australian schools, as determined by the Index of Community Socio-Educational Advantage (ICSEA). This index was developed by the Australian Curriculum, Assessment and Reporting Authority (ACARA) to provide meaningful comparisons

across Australian schools, with the national mean set at 1000 ($SD = 100$) (ACARA, 2015b). The mean demographic index for the participating schools in sample one was 1022 and the school indices ranged from 996 to 1076. In teaching spelling in these school contexts, all teachers were required to follow the Australian Curriculum (Australian, Curriculum, Assessment & Reporting Authority, ACARA, 2015a). As illustrated in "Appendix 1", teaching spelling in accordance with the national curriculum requires phonological, orthographical and morphological instruction across each grade. The participants in sample one included 381 students (178 boys and 203 girls) from Grades 3, 4, 5 and 6, aged between 8 and 12 years (see Table 4). All students whose parents provided consent were included in the sample. No participating students were diagnosed with a language or cognitive impairment; five participating students were identified as Aboriginal or Torres Strait Islander (Year 3, $n = 1$; Year 4, $n = 1$; Year 5, $n = 1$; Year 6, $n = 2$); and seven students were learning English as an additional language (Year 3, $n = 2$; Year 4, $n = 3$; Year 5, $n = 1$; Year 6, $n = 1$).

Sample two

Students for sample two were recruited from four Government schools to participate in the second phase for instrument validation. The mean demographic (ICSEA) index for the participating schools was 1028 and the school indices ranged from 985 to 1140. The participants were 457 students (228 boys and 229 girls) from Grades 3, 4, 5 and 6, aged between 8 and 12 years (see Table 5). All students whose parents provided consent were included in the sample. Eight participating students were identified as Aboriginal or Torres Strait Islander (Year 3, $n = 2$; Year 4, $n = 2$; Year 5, $n = 3$; Year 6, $n = 1$); and ten students were learning English as an additional language (Year 3, $n = 3$; Year 4, $n = 3$; Year 5, $n = 2$; Year 6, $n = 2$).

Instrument administration and scoring

Testing took place in school classrooms during the second half of the school year. For consistency, the first researcher administered and scored all tests using prescriptive scoring templates. Participating students in sample one first completed

Table 5 Demographics of participants

Grade	Boys	Girls	Mean age in years
Sample one			
Grade 3 ($n = 94$)	42	52	8
Grade 4 ($n = 99$)	44	55	9
Grade 5 ($n = 101$)	51	50	11
Grade 6 ($n = 87$)	41	46	12
Sample two			
Grade 3 ($n = 110$)	53	72	8
Grade 4 ($n = 114$)	59	55	9
Grade 5 ($n = 110$)	54	56	11
Grade 6 ($n = 123$)	62	61	12

the *CoST: Real-word version* (Daffern et al., 2015, 2017b), followed by the original *CoST: Pseudo-word version* after a short rest period. Sample two students completed the revised *CoST: Pseudo-word version* only.

Components of spelling test (CoST): real-word version

This instrument (Daffern, 2017; Daffern et al., 2015, 2017b) required students (in sample one, phase one) to spell 70 words which were presented to them orally, each within the context of a sentence. The duration for this testing was approximately 20 min. Across the 70 words, the measure comprises 101 individual items across three scales: (1) Phonological Component; (2) Orthographic Component; and (3) Morphological Component. Prescriptive scoring templates (Daffern et al., 2017b) were used to score and categorize spelling errors according to their respective spelling components. The correct spelling of an item was given a score of 1 mark while incorrect spelling was marked as 0 across the instrument.

Components of spelling test (CoST): pseudo-word version

All students in sample one were required to spell 124 newly designed pseudo-words. The pseudo-words were dictated to the students and they had to write the words on a response sheet. Note that for the morphological scale the items were presented in a cloze test form whereby students saw all of the words in the sentences written down, except for the target pseudo-word. The data collected were used for two purposes: (1) to gauge the difficulty of the items across grades; and (2) to reduce the number of initial items which amounted to 124. The analysis of the data collected from the pseudo-word instrument motivated us to design two versions of the test: *Pseudo-word-G-3-4* (for Grades 3/4) and *Pseudo-word-G-5-6* (for Grades 5/6). In the second phase of the study, students in sample two were required to complete the revised/shorter version of this instrument for validation purposes; those in Grades 3 and 4 completed the *Pseudo-word-G-3-4* while students in Grades 5 and 6 completed the *Pseudo-word-G-5-6*.

For test administration, short breaks were provided between each component of the pseudo-word test. For consistency, all items were dictated under specified timed conditions (no more than a 15 s wait time for each item to be written). Detailed scoring templates were developed for all items in each of the three scales, and responses were analyzed to identify phonological, orthographic or morphological errors. Like the real-word version of the *CoST*, the correct spelling of an item was given a score of 1 while incorrect spelling was scored 0.

Reliability and validity analyses

Construct validity

The design of the instrument was informed by current literature regarding how children learn to spell and on the linguistic structures that underpin Standard English spelling (Daffern, 2017; Treiman, 2017a). The structure of the *CoST: Pseudo-word*

version is similar to the real-word version as it contains three scales and these align with the three spelling components underpinning Triple Word Form Theory (Bahr, 2015; Garcia et al., 2010). Moreover, the instrument utilizes well-established error-analysis techniques (see, for example, Bear et al., 2012).

Content validity

Six linguistic experts (Muijs, 2004) as well as four experienced classroom educators with postgraduate qualifications (specialising in language education and inclusive or special education) were consulted to assess the linguistic suitability of each item developed for the original pseudo-word instrument. As recommended by Sireci and Falkner-Bond (2014), 10 Subject Matter Experts (SME) were requested to rate the 124 items. The SME's were asked to determine if each item reflected the linguistic feature that it was intended to measure. In doing so, they were required to assess each item on a 4-point scale (1: not relevant; 2: somewhat relevant; 3: quite relevant; 4: highly relevant). If an item was to be rated less than three, the SME's were instructed to note their reason or to suggest an alternative. The Fleiss Kappa, an index of content validity was computed for each of the three constructs. Results from the expert review process indicated that the overall agreement values for the three scales were within the acceptable 80% inter-rater agreement: (phonological scale: Fleiss Kappa = 0.93; orthographic scale: Fleiss Kappa = 0.97; and morphological scale: Fleiss Kappa = 0.97).

Inter-rater reliability for item scoring

The first author rated all the items in the first instance. As a measure of inter-rater reliability in the marking, two independent markers rated the scripts of a sample of 30 students for each grade level from sample one. Both markers had more than fifteen years of classroom teaching experience and were qualified with a postgraduate teaching qualification. One-way Fleiss Kappa was computed for each of the three constructs. In all the cases, the inter-rater agreement was almost perfect (Kottner et al., 2011), with interclass confidence interval (.999, 1). The few cases of discrepancies between markers arose as a result of the unclear handwriting of some of the students, where specific handwritten letters (<a>, <u> and <o>) were misread. The scoring of the scripts by two independent markers also pointed out the suitability of the marking scheme developed.

Descriptives and MANOVAs

For the first phase of data collection and analysis, data from sample one were used to conduct descriptive analyses, followed by a comparison of the performances of respondents across grades and linguistic components. A set of multivariate analysis of variance (MANOVA) was conducted to identify performance differences.

Predictive validity

Using data from sample one, predictive validity of the instrument was established by examining correlations of the phonological, orthographic and morphological components of the real-world version of the instrument (Daffern et al., 2015) and the newly-designed, pseudo-word version. Predictive validity analyses were conducted only for the students in sample one as they completed both the real-word CoST and pseudo-word CoST.

Item-level reliability analyses

Item-level analyses were performed during both phases of the study to gauge how the items performed from a psychometric perspective using Classical Test Theory (CTT) and Rasch. CTT and Rasch provide reliability and validity measures that are conventionally reported in psychological test calibration (Hambleton & Jones, 1993). CTT provides measures of internal consistency and corrected item-total correlation on the basis of the sample information. Rasch provides a complementary in-depth appraisal of the scale with the advantage that it is sample independent in that it takes both item difficulty and respondents' ability into consideration. Data for this study were analysed in R (version 3.2.3) and SPSS (version 22.0). For both phases, the internal consistency was established through Cronbach alpha in Classical Test Theory. Additionally, person separation reliability was computed in Rasch analysis to indicate the extent to which each sample was able to separate the items.

The results of CTT and Rasch analyses from sample one data were used to evaluate and refine the original instrument through an item reduction process. The following criteria were used to reduce the number of items: (1) Item difficulty (difficulty index < 0.2, i.e., items that were within the reach of less than 20% of the respondents were considered as inaccessible); (2) Discrimination index (discrimination index < 0.1: poor item; discrimination index between 0.1 and 0.3: fair discrimination; discrimination index > 0.3: good discrimination); and (3) Misfit items (Items outside the range $-2 <$ Standardised fit statistic < 2 and $0.5 <$ Mean Square fit statistic < 1.5 (Linacre, 2002) across Grades 3/4 and Grades 5/6 were considered misfits).

Data from sample two were then used to compute descriptive statistics, item difficulty, item discrimination, infit statistics, outfit statistics and separation reliability on *Pseudo-word-G-3-4* and *Pseudo-word-G-5-6* (the revised instruments).

Results

Descriptives: sample one

Table 6 provides descriptive statistics for each of the three components of the initial iteration of the instrument. It can be observed that the mean scores for each of the components increase across age, reflective of the growing maturity of the participants from Grade 3 to Grade 6. Results from the Phonological Component indicate that as the number of syllables increases from monosyllabic to polysyllabic,

Table 6 Mean and standard deviation of the three scales in the initial instrument

	n Items	Grade 3 Mean (SD)	Grade 4 Mean (SD)	Grade 5 Mean (SD)	Grade 6 Mean (SD)
Phonological component					
Monosyllabic	10	7.28 (2.73)	7.90 (1.84)	8.33 (1.58)	8.77 (1.44)
Disyllabic	10	4.57 (3.03)	4.89 (2.41)	5.61 (2.39)	6.63 (2.77)
Polysyllabic	10	1.94 (2.16)	2.22 (2.24)	2.79 (2.55)	4.09 (3.16)
Phonological total	30	13.79 (7.09)	15.01 (5.48)	16.73 (5.72)	19.49 (6.54)
Orthographic component					
Orthographic (Part A)	28	12.70 (6.06)	15.15 (5.36)	17.70 (4.90)	19.23 (3.67)
% items correct		45.4	54.1	63.2	68.7
Orthographic (Part B)	11	6.67 (2.77)	7.17 (2.40)	8.53 (1.89)	9.34 (1.40)
% items correct		60.6	65.2	77.5	84.9
Orthographic total	39	19.32 (8.15)	22.32 (7.03)	26.24 (6.31)	28.57 (4.66)
Morphological component					
Inflected suffix	23	10.61 (4.24)	11.23 (4.28)	14.41 (4.65)	18.11 (3.72)
% items correct		46.1	48.8	62.7	78.7
Derivational suffix	16	5.47 (3.58)	6.81 (3.27)	8.28 (3.11)	11.33 (2.74)
% items correct		34.2	42.6	51.8	70.8
Prefixes	8	3.90 (1.45)	4.36 (1.20)	5.06 (1.46)	5.80 (1.55)
% items correct		48.8	54.5	63.3	72.5
Roots	8	2.01 (2.36)	2.79 (2.35)	3.68 (2.41)	5.56 (1.99)
% items correct		25.1	34.9	46.0	69.5
Morphological total	55	21.99 (9.91)	25.19 (9.33)	31.43 (10.0)	40.82 (8.45)

As the number of items is different in the constructs for orthographic and morphological components, percentages are provided to facilitate comparison

mean scores decrease. In the Orthographic Component, performance was higher in Part B (common long vowel patterns) than that in Part A (constrained letter pattern) across grade levels. In the Morphological Component, performance was higher in the pseudo-words involving prefixes (with the exception of the Grade 6 students) and the scores were lowest in the root words.

Grade-level performance comparisons (sample one)

Multivariate analysis of variance was conducted with the phonological, orthographic and morphological scores as dependent variables and the four grade levels as independent variables (see Table 7) to identify significant differences across grades. There was a significant effect of grade level on the phonological, orthographic and morphological scores (Pillai's trace, $v = 0.424, F(9, 1131) = 20.67, p = .001$). The post hoc Bonferroni results are presented in Table 7.

The post hoc Bonferroni test (Table 7) shows that differences were significant ($p < 0.01$) in the pairwise comparisons with the following exceptions: (1) between Grades 3 and Grade 4 students in the Phonological and Morphological Components,

Table 7 Differences in phonological, orthographic and morphological scores across successive grades

Dependent variables	Bonferroni post hoc test (p values)		
	G3/G4	G4/G5	G5/G6
Phonology $F(3, 377) = 14.32, p < .01$	1.000	.305	.015*
Orthography $F(3, 377) = 39.15, p < .01$	.014*	.001**	.104
Morphology $F(3, 377) = 69.33, p < .01$	.116	.001**	.001**

G3/G4 refers to comparison of scores in Grades 3 and 4

*$p < .05$, **$p < .01$

Table 8 Differences in monosyllabic, disyllabic and polysyllabic scores across successive grades

Dependent variables	Bonferroni post hoc test (p values)		
	G3/G4	G4/G5	G5/G6
Monosyllabic $F(3, 377) = 9.62, p < .01$	.170	.746	.741
Disyllabic $F(3, 377) = 10.74, p < .01$	1.000	.324	.054
Polysyllabic $F(3, 377) = 12.73, p < .01$	1.000	.682	.003**

G3/G4 refers to comparison of scores in Grades 3 and 4

Pillai's trace, $v = .121$, $F(9, 1131) = 5.30$, $p = .001$

*$p < .05$, **$p < .01$

(2) between Grade 4 and Grade 5 students in the Phonological Component, and (3) between Grade 5 and Grade 6 students in the Orthographic Component.

As each of the three scales were developed on the basis of constructs (or sub-skills), we performed further comparisons to observe how these constructs varied across Grade levels. Tables 8, 9 and 10 show the MANOVA results for the constructs. In the Phonological Component (Table 8), differences were insignificant, except in Grades 5 and 6 in the polysyllabic construct. In the Orthographic

Table 9 Differences in scores in constrained and common long vowel patterns across successive grades

Dependent variables	Bonferroni post hoc test (p values)		
	G3/G4	G4/G5	G5/G6
Part A (Constrained letter pattern) $F(3, 377) = 29.34$, $p < .01$	.006**	.003**	.247
Part B (Common long vowel patterns) $F(3, 377) = 38.25$, $p < .01$	.673	.001**	.070**

Pillai's trace, $v = 0.270$, $F(6, 754) = 19.59$, $p = .001$

*$p < .05$, **$p < .01$

Table 10 Differences in scores in inflected suffix, derivational suffix, prefix and root across successive grades

Dependent variables	Bonferroni post hoc test (p values)		
	G3/G4	G4/G5	G5/G6
Inflected suffix $F(3, 377) = 59.58, p < .01$	1.000	.001**	.001**
Derivational suffix $F(3, 377) = 55.91, p < .01$	.023*	.008**	.001**
Prefix $F(3, 377) = 31.30, p < .00$	.149	.003**	.002**
Roots $F(3, 377) = 40.14, p < .01$	.114	.036*	.001**

Pillai's trace, $v = .387$, $F(12, 1128) = 13.92$, $p = .001$

*$p < .05$, **$p < .01$

Component (Table 9), differences are significant except in the following two cases: (1) Grade 5 and Grade 6 (Part A) and (3) Grade 3 and 4 (Part B). In the Morphological Component (Table 10), differences are significant except primarily between the Grade 3 and Grade 4 students.

Item reduction

In developing the instrument, a large number of items were initially included in order to assess item accessibility for students across each grade, and to assess the items' suitability in measuring the intended dimensions. The intention was to produce a shorter version of the instrument with a reduced number of items that parsimoniously tap on the constructs. In the course of phase one analysis (using data from sample one), we found it compelling to use the items to develop two versions of the test: *Pseudo-word-G-3-4* (for Grades 3/4) and *Pseudo-word-G-5-6* (for Grades 5/6). This decision was made because students in Grades 3 and 4 were closer in developmental levels than those in Grades 5 and 6 as a group. Moreover, some of the items were too difficult for the Grade 3 and Grade 4 students and it would be inappropriate to test their spelling ability on such items.

CTT and Rasch results

Due to space limitations, we provide detailed CTT and Rasch results only for the Phonological Component for Grades 3 and 4 and provide the range of values for item parameters for the Orthographic and Morphological Components (see "Appendix 2", Tables 19, 20). First, we comment on the CTT results. The item difficulties for the Phonological Component across the four grade levels ranged from .03 to .95 and the discrimination indices ranged from .01 to .69. Similarly, the item difficulties for the Orthographic Component for the four grade levels ranged from .12 to .99 and the discrimination indices ranged from − 0.20 to .68. The item difficulties for the

Morphological Component for the four grade levels varied from .01 to .99 while the discrimination indices ranged from – 0.21 to 0.66.

In the Rasch analysis, item difficulties and outfit and infit indices were computed. The infit and outfit values for the three linguistic components across the four grade levels varied as follows: (1) Phonological: − 3.97 to 3.24; (2) Orthographic: − 3.93 to 4.39; (3) Morphological: − 3.65 to 4.33; although the majority of the values were in the range − 2 to 2. The application of the criteria described in the method section led to the reduction of items for each of the scales to produce the two final versions of the instrument. We also adjusted the final number of items so that they rounded up to the nearest 5 or 10. For example, the Phonological Component of the *Pseudo-word-G-3-4* contained 24 items after the application of the reduction criteria. We added one more item (taken from the initial version) which minimally affected the psychometric properties to get a 25-item instrument.

Descriptives: sample two

The descriptive statistics for the reduced/final version of the instrument is presented in Table 11. As expected, the mean value of the three spelling components for Grades 5 and 6 is greater than that of Grades 3 and 4.

Reliability of the instrument

Table 12 presents the reliability values of the revised instrument (*Pseudo-word-G-3-4* and *Pseudo-word-G-5-6*). It shows strong internal consistency among the items in each of the constructs. The Cronbach alpha values range from .812 to .931 and the separation reliability values vary from .790 to .916, well above the 0.7 recommended benchmark.

Predictive validity

The pseudo-word instrument was found to significantly correlate with the real-word version of the *CoST* (see Table 13). The numbers below the diagonal are the correlations between the real-word and pseudo-word constructs for the Grade 3 and 4 students (taken as one cohort) while those above the diagonal are for Grade 5 and 6

Table 11 Mean and standard deviation for the reduced instrument

Spelling component	Pseudo-word-G-3-4		Pseudo-word-G-5-6	
	($n = 224$)	($n = 233$)	$n = 30$	
	No. of items	Mean (SD)	No. of items	Mean (SD)
Phonological	25	13.70 (5.36)	$n = 30$	18.16 (6.33)
Orthographic	30	16.94 (6.02)	$n = 35$	24.39 (5.39)
Morphological	45	21.88 (8.67)	$n = 50$	33.79 (9.84)

Table 12 Reliability indices (Cronbach alpha and separation reliability) of the reduced instrument

Students	$n=224$		$n=224$	
CoST	Pseudo-word-G-3-4		Pseudo-word-G-5-6	
	Items	Cronbach alpha (separation reliability)	Items	Cronbach alpha (separation reliability)
Phonological	$n=25$	$\alpha=.865$ (.860)	$n=30$	$\alpha=.897$ (.909)
Orthographic	$n=30$	$\alpha=.854$ (.845)	$n=35$	$\alpha=.812$ (.790)
Morphological	$n=45$	$\alpha=.910$ (.913)	$n=50$	$\alpha=.931$ (.916)

Separation reliability indices from Rasch are presented in brackets

Table 13 Correlation between *CoST: Real-word* and *CoST: Pseudo-word* for Grades 3–4 and Grades 5–6

Spelling component	1	2	3	4	5	6
1. Phonological-rw	–	.796**	.785**	.681**	.702**	.770**
2. Orthographic-rw	.809**	–	.826**	.572 **	.775**	.833**
3. Morphological-rw	.811**	.860**	–	.669**	.761**	.883**
4. Phonological-pw	.806**	.749**	.733**	–	.713**	.614**
5. Orthographic-pw	.781**	.800**	.757**	.804**	–	.717**
6. Morphological-pw	.793**	.820**	.851**	.770**	.777**	–

rw real word, *pw* pseudo-word

**$p<.01$

students (taken as one cohort). For instance, the correlation between Phonological-rw and Orthographic-rw is .796 for the Grades 5 and 6 cohort while it is .809 for the Grade 3 and 4 cohort. The correlations are significant and relatively high for both the Grade 3 and 4 cohort and the Grade 5 and 6 cohort, supporting the argument that the spelling components develop almost concurrently across age level.

To further support the claim that the three spelling skills broadly develop concurrently rather than in stages, we used the net scores in the instrument (i.e., the sum scores for phonology, orthography and morphology) to categorize the Grades 3–4 and Grades 5–6 students as low, medium and high (see Table 14) based on the quartile (25%, 50% and 75%) scores. Then, the mean scores for the individual scales (phonology, orthography and morphology) in the low, medium and high groups were computed for Grades 3–4 and Grades 5–6. It can be observed that there is much consistency in performance across the three constructs for the low, medium and high groups at both Grades 3–4 and Grades 5–6. In other words, if a student has a low score in the phonological scale, then they also have a low score in the orthographic or morphological scale. Thus, these data support the fact that the three skills develop concurrently. We also computed the differences among the phonological, orthographic and morphological skills for the low, medium and high performers separately and significant differences were observed. Thus, although the three scales are related, they are also distinct to some extent.

Table 14 Mean (and standard deviation) differentiated by constructs, level of students and Grade levels

	Low		Medium		High	
	Grade 3–4 ($n=59$)	Grade 5–6 ($n=60$)	Grade 3–4 ($n=112$)	Grade 5–6 ($n=115$)	Grade 3–4 ($n=53$)	Grade 5–6 ($n=58$)
Phonology	7.46 (3.34)	11.07 (3.67)	12.40 (2.61)	17.99 (3.88)	17.69 (3.37)	25.84 (2.49)
Orthography	7.94 (2.98)	15.29 (3.53)	12.91 (2.39)	21.73 (2.99)	18.04 (2.66)	25.11 (1.91)
Morphology	6.49 (2.58)	13.13 (4.71)	10.64 (1.65)	20.77 (3.00)	15.95 (3.07)	26.68 (2.21)

Table 15 Comparison of performance between a sample of two real-word and pseudo-word inflected suffixes

Inflected suffixes	Generalization 1: If a base word ends with a short vowel grapheme followed by a single consonant grapheme, double the final consonant and add the suffix marker (e.g. '-ed' for past tense)		Generalization 2: If a base word ends in the grapheme 'e', drop the 'e' then add the tense suffix marker '-ing' or '-ed'	
	Real-word item *knot-knotted*	Pseudo-word item *clom-clommed*	Real-word item *serve-serving*	Pseudo-word item *vipe-viping*
Grade 3–4	89.1	6.2	99.5	30.9
Grade 5–6	87.0	63.7	94.1	82.4

It should be highlighted that the two tests are not interchangeable but should be used complementarily. For example, the real-word test may be taken by a group of students in Grade 3 and then in Grade 5 while the same students may take the pseudo-word test when they are in Grade 4 and then when they proceed to Grade 6. This proposed schedule of testing will help to minimise threats to validity associated with any retesting, and it enables teachers, clinicians or researchers to use the data from both tests for diagnostic and summative assessment purposes. The real word version of the spelling test is based on contextualised words that students may have heard or come across from their schooling or out of schooling experiences. On the other hand, the pseudo-word version provides a measure of spelling knowledge devoid of word-specific knowledge and as such it tests if a student is able to apply underlying plausible spelling generalizations to unknown words. We compared students' ability to spell the pseudo-words and parallel real words to illustrate the possible influence of word-specific contextual knowledge. To illustrate, Table 15 provides an example using two inflected suffix items from the real-word test and two correspondingly parallel items from the pseudo-word test in terms of their linguistic properties. The results demonstrate that both Grade 3–4 and Grade 5–6 students performed better in inflected suffixes involving real words than the linguistically parallel pseudo-word items.

Test norms

To inform potential school-based intervention plans, data from sample two were used to develop test norms based on percentiles. Tables 16 and 17 respectively show the distribution of students at the 5th, 10th, 25th, 50th, 75th, 90th and 95th percentiles for *Pseudo-word-G-3-4* and *Pseudo-word-G-5-6*.

Table 16 Test norms for *Pseudo-word-G-3-4*

Percentile	Phonological score	Orthographic score	Morphological score
5	5	6	8
10	6	8	11
25	10	12	16
50	14	18	21
75	18	22	29
90	20	24	34
95	22	26	35

Table 17 Test norms for *Pseudo-word-G-5-6*

Percentile	Phonological score	Orthographic score	Morphological score
5	7	13	15
10	9	18	18
25	13	21	27
50	18	25	36
75	24	29	40
90	26	31	46
95	28	32	48

Discussion

The primary purpose of this study was to design and validate a new dictation test, labelled the *Components of Spelling Test (CoST): Pseudo-word version*. The *CoST: Pseudo-word version* is a measure of spelling ability for students in the age range of 8 to 12 years. It has been designed to help teachers effectively plan for spelling instruction in school contexts and for specialist clinicians to deliver suitable interventions for students experiencing difficulties with spelling. This instrument has also been designed as a measure of spelling ability for experimental and correlational research purposes. This novel instrument fills a gap in spelling ability research literature by providing the first pseudo-word metric to assess 8- to 12-year-old students' phonological, orthographic and morphological spelling skills.

For the first (calibration) phase of the study, we sought to determine the extent to which the phonological, orthographic and morphological items in the initial iteration of the instrument were accessible to students in Grades 3–6 (aged 8–12 years). Consistent with an earlier study involving student performance data from the real-word version of the *CoST* (Daffern, 2017), scores across each component increased as a function of grade.

In the Phonological Component, students achieved higher scores in the Monosyllabic and Disyllabic items than the Polysyllabic items across grades. These results suggest that spelling errors are more likely to occur when a larger number of consecutive

phonemes need to be retained in working memory, analysed and then encoded in their correct order. Indeed, the observed reduction of scores in the Polysyllabic construct resonates with previous research demonstrating the role of phonological working memory in spelling and reading, whereby cognitive load is decreased if fewer consecutive phonemes and their corresponding graphemes need to be encoded (Daffern, 2017; Gathercole, 2007; Gathercole & Baddeley, 1993; Plaza & Cohen, 2003).

In the Orthographic Component, the results indicate that scores for the constrained letter pattern (Part A) were lower than they were for the common long vowel pattern (Part B) and grade level differences were significant, with the exception of Grade 5 to Grade 6 in Part A and Grades 3 to Grades 4 in Part B. One possible explanation is that each item measured in Part B can be spelled in multiple plausible ways (for example, <blate>, <blait>, <blaight> or <bleight>), whereas each item in Part A only has one plausible response (e.g., the only way to spell the diphthong in the word 'zoy' is <-oy>, not <-oi>). These results suggest that students may learn plausible alternations for common long vowels with relative ease. Orthographic knowledge, at least to some extent, may be a function of word specific knowledge, partly developed over time through exposure to print (Graham, 2000; Treiman, 2018). For example, even though a test-taker may know that the letter patterns <oo>, <oe>, <ough> can plausibly represent the same phoneme, the test-taker may not necessarily know which of these choices is appropriate when applying it to a specific real word (for example, spelling <sm**oo**th> with double <o> is correct but <sm**oe**th> or <sm**ough**th> is incorrect). Consequently, a pseudo-word measure may need to be accompanied by a real-word measure in order to make an adequate judgment of a student's spelling ability. Easier items in the constrained letter pattern (Part A) were those which required the student to know, for example, that it is not plausible to start a word with the letters <ck>, or that the letters <oi> never appear at the end of a base word, or that the letter <e> always follows the letter <v> in a base word. Teachers or clinicians could utilize such insights to decide which orthographic patterns individual students are yet to master.

In the Morphological Component, scores in Inflected Suffixes were higher than Derivational Suffixes across grades; and students scored lowest in the Greek and Latin Root construct across grades. These results parallel the findings obtained from the real-word instrument (Daffern, 2017). It is further noted that root items (e.g., <psych>and <chrono>) were barely accessible to the Grade 3 students. This was anticipated considering etymology (the study of word origins) is not typically addressed until the later primary school grades in Australia (ACARA, 2015a).

While the first phase of the study ensured that the instrument inherited content and construct validity through an expert review process and a well-established theoretical foundation, the second phase established the internal consistency of the instrument, with Cronbach alpha and separation reliability values showing strong internal consistency among the items in each of the two versions of the instrument. This study has also shown that students who do well in real-word spelling as measured by the real-word *CoST* tend to do well in pseudo-word spelling, although the two instruments vary in the underlying sub-constructs (Table 1). Likewise, students who perform poorly in the real-word test tend to perform poorly in the pseudo-word test. Even though high correlations have been observed between the real-word and

pseudo-word tests, there is utility in using both instruments interchangeably. Specifically, problems can arise if the real-word test is used many times with the same group of students. The introduction of a new test with different items is helpful for longer term utility in tracking growth and for providing ongoing support to students across Grades 3–6. Furthermore, ceiling effects in the Orthographic Component of the real-word test have been reported in another study (Daffern, 2017), yet no ceiling was observed in the Orthographic Component of the pseudo-word test. Thus, despite the high correlations between the real and pseudo-word tests, there is greater scope for assessment using the pseudo-word test with high performing spellers due to the inclusion of more difficult items.

Applications of the new instrument

As Treiman (2018) contends, in order to support student learning in spelling, there is a need for educators to understand how the written language system works and to have the skills and resources to be able to identify and interpret the errors that students make. This new instrument provides both school-based educators and researchers a robust tool to be able to measure student learning in spelling. Teachers or clinicians may use this tool for diagnostic purposes, and to evaluate the effectiveness of their own teaching. Researchers may use this tool to obtain fine-grained understandings of how children learn to spell across a range of student populations, or to measure the efficacy of interventions seeking to improve outcomes in spelling ability. To further assist educators and researchers in using this new instrument, instructional and administrative recommendations have since been developed to accompany the constructs within. Educators and researchers may contact the first author (via http://www.tessadaffern.com) to request access to the instrument and the accompanying instructional recommendations.

Conclusion

The results of the present study contribute to the literature in a number of ways: First, a new measure of spelling is offered to help classroom teachers and clinicians accurately identify a respondent's knowledge of linguistic generalisations in spelling without the dependence on word-specific knowledge. The new instrument is freely available at http://www.tessadaffern.com. Second, we present the first pseudo-word instrument informed by Triple Word Form Theory, which is premised on the assumption that linguistic skills in spelling can develop concurrently rather than in sequential stages. Current spelling instruments (e.g., the *Words Their Way Inventories*) are based on stage theory. Third, given that the instrument is based on specific linguistic features, it provides a clear indication where respondents may be lagging behind in spelling. That is, it can be used to perform a spelling error analysis (Al Otaiba & Hosp, 2010). Fourth, from a practical perspective, it provides a comprehensive range of items to measure linguistic skills across three overarching components of spelling, thus minimising the need for a teacher or clinician to administer other assessments. Finally, compared to existing pseudo-word instruments (Kohnen et al., 2015), this

new instrument includes phonological complexities (not only monosyllabic but also disyllabic and polysyllabic word structures) and morphological complexities (that is, a large range of inflected and derivational suffixes, prefixes, and Greek and Latin root structures.

In developing and validating this instrument we sought to reduce a gap in spelling ability research literature by providing the first pseudo-word metric to assess 8- to 12-year-old students' phonological, orthographic and morphological spelling skills in one instrument. Compared with widely used dictation spelling tests such as the *Words Their Way Inventories* (Bear et al., 2012) and the *South Australian Spelling Test* (Westwood, 2005), the *CoST: Pseudo-word version* assesses a comprehensive range of spelling subskills. The test is user-friendly as it can be administered to a whole class of students at one time or to an individual student within approximately 40 min. Prescriptive scoring templates are also available for each component, making it an easy and efficient tool to use.

While the instrument exhibits robust psychometric properties, this form of spelling assessment should not necessarily replace another. A combination of both real-word and pseudo-word spelling measures is of value to educators developing intervention plans (Kohnen et al., 2009). In determining the full scope of strengths and weaknesses in a student's spelling ability, insights from the *CoST: Pseudo-word version* may also be complemented by qualitative analyses of the spelling errors a student makes in the context of freely composed writing (see, for example, Daffern, 2016).

Limitations and future directions

All test items in this novel instrument were carefully developed to ensure their linguistic relevance across the various Standard Englishes (e.g., Australian, British, American, NZ and Canadian English). As this new test measures accuracy in spelling linguistic generalisations in Standard English, the results are not expected to be very different in other countries where English is the mother tongue. Further, while the results are unlikely to differ in other Australian states, as the mean ICSEA for this study is similar to the national mean, we welcome further testing with older student populations or where spelling curricula may be substantially different. Although norms are not yet developed for populations of students beyond an Australian context, this newly developed instrument is useful when administered without reference to any norms because it is a comprehensive assessment of the most relevant spelling skills. Thus, if a construct reveals gaps, the test data can inform teachers or clinicians to target the relevant skills not currently mastered by a student or client.

The construction of this new instrument may inform future developments of spelling ability instruments in languages other than English. There is scope to conduct additional validation testing by, for example, establishing the test–retest reliability and stability of this measure over time. A future study should also examine if performance in the three pseudo-word measures predict spelling and reading performance in later grades, beyond performance in real-word spelling. Furthermore, to expand the utility of this instrument, the development of norms, based on respondents of other student populations is recommended. Research is currently underway to design and validate a *CoST* for the early

769 school years, whereby phonological, orthographic and morphological scales are con-
770 structed for testing among students aged 5–7 years.

771 Appendix 1

772 See Table 18.

Table 18 Teaching spelling (Grades 3–6) in accordance with the Australian curriculum: English

Australian curriculum: English spelling content descriptors[a]	Dominant teaching focus
Grade 3	
Understand how to apply knowledge of letter-sound relationships, syllables, and blending and segmenting to fluently read and write multisyllabic words with more complex letter patterns	P
Recognise and know how to write most high-frequency words including some homophones	O & M
Understand how to use letter-sound relationships and less common letter patterns to spell words	P & O
Grade 4	
Understand how to use knowledge of letter patterns including double letters, spelling generalisations, morphemic word families, common prefixes and suffixes and word origins to spell more complex words	O & M
Understand how to use knowledge of letter patterns including double letters, spelling generalisations, morphemic word families, common prefixes and suffixes and word origins to spell more complex words	M
Read and write a large core of high frequency words including homophones and know how to use context to identify correct spelling	O & M
Understand how to use phonic knowledge to read and write multisyllabic words with more complex letter combinations, including a variety of vowel sounds and known prefixes and suffixes	P, O, & M
Grade 5	
Explore less common plurals, and understand how a suffix changes the meaning or grammatical form of a word	M
Understand how to use phonic knowledge to read and write less familiar words that share common letter patterns but have different pronunciations	P & O
Understand how to use knowledge of known words, base words, prefixes and suffixes, word origins, letter patterns and spelling generalisations to spell new words	M & O
Grade 6	
Understand how to use phonic knowledge and accumulated understandings about blending, letter-sound relationships, common and uncommon letter patterns and phonic generalisations to read and write increasingly complex words	P & O
Understand how to use knowledge of known words, word origins including some Latin and Greek roots, base words, prefixes, suffixes, letter patterns and spelling generalisations to spell new words including technical words	M & O

P phonological, *O* orthographic, *M* morphological

[a]ACARA (2016)

Appendix 2

See Tables 19 and 20.

Table 19 Item difficulty, discrimination and fit statistics of the phonological component for Grades 3 and 4 (initial instrument)

Item	Grade 3					Grade 4				
	CTT diff.	CTT discr.	Rasch diff.	Outfit t	Infit t	CTT diff.	CTT discr.	Rasch diff.	Outfit t	Infit t
pwphon1	.77	0.50	2.15	0.31	0.16	.87	0.27	2.40	−0.10	0.49
pwphon2	.79	0.49	2.34	−0.29	0.56	.86	0.33	2.30	−0.09	−0.28
pwphon3	.74	0.58	1.97	−0.55	−1.00	.74	0.33	1.35	0.26	0.65
pwphon4	.81	0.40	2.54	0.66	1.09	.84	0.26	2.11	0.38	0.39
pwphon5	.54	0.57	0.57	−0.18	−0.40	.71	0.51	1.16	−1.05	−1.23
pwphon6	.71	0.39	1.72	1.22	1.48	.73	0.24	1.28	0.95	1.53
pwphon7	.54	0.60	0.57	−0.61	−0.80	.46	0.39	−0.16	1.09	0.34
pwphon8	.86	0.43	3.14	0.79	−0.45	.92	0.13	3.02	0.87	0.54
pwphon9	.80	0.54	2.44	−0.48	−0.37	.94	0.39	3.36	−1.05	−0.81
pwphon10	.71	0.55	1.72	−0.43	−0.11	.84	0.30	2.11	1.56	−0.04
pwphon11	.67	0.55	1.41	−0.49	0.18	.73	0.44	1.28	−0.73	−0.37
pwphon12	.55	0.61	0.64	−0.96	−0.82	.66	0.23	0.86	1.21	2.14
pwphon13	.38	0.50	−0.41	0.64	0.44	.33	0.43	−0.86	−0.56	−0.19
pwphon14	.36	0.63	−0.55	−1.21	−1.73	.44	0.53	−0.26	−1.56	−1.47
pwphon15	.59	0.67	0.84	−1.43	−1.89	.71	0.43	1.16	1.31	−0.48
pwphon16	.54	0.58	0.57	−0.65	−0.29	.55	0.43	0.26	−0.49	0.08
pwphon17	.38	0.56	−0.41	−0.68	−0.35	.34	0.51	−0.80	−1.42	−1.24
pwphon18	.46	0.49	0.05	0.76	0.77	.44	0.36	−0.26	0.78	0.88
pwphon19	.10	0.40	−2.86	−0.22	−0.67	.03	0.10	−4.10	0.37	0.24
pwphon20	.54	0.54	0.57	−0.07	0.21	.66	0.58	0.86	−1.74	−2.04
pwphon21	.10	0.31	−2.86	0.44	0.05	.14	0.31	−2.24	0.23	−0.42
pwphon22	.20	0.37	−1.70	0.35	0.89	.24	0.32	−1.42	0.01	0.76
pwphon23	.30	0.55	−0.97	−0.42	−0.69	.32	0.52	−0.92	−1.52	−1.58
pwphon24	.15	0.43	−2.21	−0.20	−0.25	.19	0.48	−1.79	−1.24	−1.27
pwphon25	.21	0.30	−1.61	1.30	2.07	.24	0.45	−1.42	−0.69	−0.84
pwphon26	.27	0.46	−1.20	0.03	0.44	.23	0.43	−1.49	−0.88	−0.58
pwphon27	.17	0.49	−1.99	−0.50	−0.89	.15	0.34	−2.14	−0.41	0.18
pwphon28	.35	0.52	−0.62	−0.25	−0.02	.48	0.42	−0.06	−0.06	0.08
pwphon29	.06	0.42	−3.40	−0.41	−1.23	.05	0.37	−3.53	−0.98	−0.74
pwphon30	.13	0.44	−2.44	−0.19	−0.89	.16	0.30	−2.05	−0.04	0.17

CTT diff item difficulty based on CTT, *CTT discr* item discrimination based on CTT, *Rasch diff* item difficulty based on Rasch

Table 20 Item difficulty, discrimination and fit statistics of the phonological component in *Pseudo-word-G-3-4* (final instrument)

Item	CTT diff.	CTT discr.	Rasch diff.	Outfit t	Infit t
pwphon1	0.83	0.34	1.80	0.07	0.33
pwphon2	0.83	0.36	1.76	−0.49	0.42
pwphon3	0.76	0.40	1.23	0.05	−0.09
pwphon4	0.85	0.19	1.93	1.31	2.05
pwphon5	0.59	0.49	0.16	−1.00	−0.88
pwphon6	0.70	0.31	0.78	1.42	1.91
pwphon7	0.50	0.40	−0.35	1.11	0.86
pwphon8	0.91	0.30	2.60	0.52	−0.76
pwphon9	0.85	0.42	1.93	−1.04	−0.68
pwphon10	0.78	0.45	1.36	0.51	−0.77
pwphon11	0.71	0.45	0.84	−1.02	−0.33
pwphon12	0.59	0.34	0.16	0.98	2.30
pwphon13	0.34	0.46	−1.18	−1.15	−0.37
pwphon14	0.41	0.57	−0.83	−2.28	−2.69
pwphon15	0.64	0.59	0.46	−1.82	−3.04
pwphon16	0.53	0.52	−0.18	−1.56	−1.25
pwphon17	0.35	0.53	−1.13	−1.85	−1.92
pwphon18	0.46	0.40	−0.51	1.26	1.06
pwphon20	0.61	0.55	0.26	−1.63	−2.23
pwphon22	0.21	0.30	−2.08	0.10	0.80
pwphon23	0.28	0.46	−1.59	−1.05	−0.91
pwphon25	0.20	0.36	−2.15	0.28	−0.11
pwphon26	0.24	0.38	−1.82	−0.13	0.00
pwphon27	0.14	0.42	−2.61	−1.27	−1.85
pwphon28	0.40	0.42	−0.85	−0.06	0.34

CTT diff item difficulty based on classical test theory, *CTT discr* item discrimination based on CTT, *Rasch diff* item difficulty based on Rasch

References

Al Otaiba, S., & Hosp, J. (2010). Spell it out: The need for detailed spelling assessment to inform instruction. *Assessment for Effective Intervention, 36*(1), 3–6.

Apel, K. (2014). A comprehensive definition of morphological awareness: Implications for assessment. *Topics in Language Disorders, 34*(3), 197–209. https://doi.org/10.1097/TLD.0000000000000019.

Australian Curriculum, Assessment, & Reporting Authority (ACARA). (2015a). Australian Curriculum English: Scope and sequence. Retrieved from http://www.australiancurriculum.edu.au/Australian%20Curriculum.pdf?Type=0&s=E&e=ScopeAndSequence.

Australian Curriculum, Assessment, & Reporting Authority (ACARA). (2015b). ICSEA: My school fact sheet. Retrieved 12 Nov, 2016. https://acaraweb.blob.core.windows.net/resources/About_icsea_2014.pdf.

Australian Curriculum, Assessment, & Reporting Authority (ACARA). (2016). NAP national assessment program. Retrieved from http://www.nap.edu.au/about/about.html.

Bahr, R. (2015). Spelling strategies and word formation processes. In R. Bahr & E. Silliman (Eds.), *Routledge handbook of communication disorders* (pp. 193–203). London: Routledge.

Bahr, R., Silliman, E., Berninger, V., & Dow, M. (2012). Linguistic pattern analysis of misspellings of typically developing writers in grades 1–9. *Journal of Speech, Language, and Hearing Research, 55,* 1587–1599.

Bear, D. R., Invernizzi, M., Templeton, S., & Johnston, F. (2012). *Words their way: Word study for phonics, vocabulary, and spelling instruction* (5th ed.). Upper Saddle River, NJ: Pearson Education Inc.

Bear, D. R., & Templeton, S. (1998). Explorations in developmental spelling: Foundations for learning and teaching phonics, spelling and vocabulary. *The Reading Teacher, 52*(3), 222–242.

Berninger, V., Vaughan, K., Abbott, R., Abbott, S., Rogan, L., Reed, E., et al. (1998). Early intervention of spelling problems: Teaching functional spelling units of varying size with a multiple-connections framework. *Journal of Educational Psychology, 90*(4), 587–605.

Bowers, J., & Bowers, P. (2017). Beyond phonics: The case for teaching children the logic of the English spelling system. *Educational Psychologist, 52*(2), 124–141. https://doi.org/10.1080/00461520.2017.1288571.

Breaux, K. C., Bray, M. A., Root, M. M., & Kaufman, A. S. (2017). Introduction to special issue and to KTEA-3 error analysis. *Journal of Psychoeducational Assessment, 35*(1–2), 4–6. https://doi.org/10.1177/0734282916669656.

Bryant, P., & Nunes, T. (2009). Morphemes and children's spelling. In R. Beard, D. Myhill, J. Riley, & M. Nystrand (Eds.), *The SAGE handbook of writing development* (pp. 329–348). London: SAGE Publications.

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

Daffern, T. (2015). Helping students become linguistic inquirers: A focus on spelling. *Literacy Learning: The Middle Years, 23*(1), 33–39.

Daffern, T. (2016). What happens when a teacher uses metalanguage to teach spelling? *The Reading Teacher, 70*(4), 423–434. https://doi.org/10.1002/trtr.1528.

Daffern, T. (2017). Linguistic skills involved in learning to spell: An Australian study. *Language and Education, 31*(1), 307–329. https://doi.org/10.1080/09500782.2017.1296855.

Daffern, T. (2018). Spelling assessment, learning, and instruction in VET. In S. McGrath, M. Mulder, J. Papier, & R. Suart (Eds.), *Handbook of vocational education and training: Developments in the changing world of work* (Vol. Section 7). Cham: Springer.

Daffern, T., Mackenzie, N. M., & Hemmings, B. (2015). The development of a spelling assessment tool informed by Triple Word Form Theory. *Australian Journal of Language and Literacy, 38*(2), 72–82.

Daffern, T., Mackenzie, N. M., & Hemmings, B. (2017a). Predictors of writing success: How important are spelling, grammar and punctuation? *Australian Journal of Education, 61*(1), 75–87. https://doi.org/10.1177/0004944116685319.

Daffern, T., Mackenzie, N. M., & Hemmings, B. (2017b). Testing spelling: How does a dictation method measure up to a proofreading and editing format? *Australian Journal of Language and Literacy, 40*(1), 28–45.

Devonshire, V., & Fluck, M. (2010). Spelling development: Fine-tuning strategy-use and capitalising on the connections between words. *Learning and Instruction, 20,* 361–371.

Devonshire, V., Morris, P., & Fluck, M. (2013). Spelling and reading development: The effect of teaching children multiple levels of representation in their orthography. *Learning and Instruction, 25,* 85–94. https://doi.org/10.1016/j.learninstruc.2012.11.007.

Ehri, L. C. (2000). Learning to read and learning to spell: Two sides of a coin. *Topics in Language Disorders, 20*(3), 19–49.

Ehri, L. C. (2005). Learning to read words: Theory, findings, and issues. *Scientific Studies of Reading, 9*(2), 167–188. https://doi.org/10.1207/s1532799xssr0902_4.

Ganske, K. (1999). The developmental spelling analysis: A measure of orthographic knowledge. *Educational Assessment, 6*(1), 41–70.

Garcia, N., Abbott, R., & Berninger, V. (2010). Predicting poor, average, and superior spellers in grades 1–6 from phonological, orthographic, and morphological, spelling, or reading composites. *Written Language and Literacy, 13*(1), 61–98.

Gathercole, S. (2007). Working memory: A system for learning. In R. K. Wagner, A. E. Muse, & K. R. Tannenbaum (Eds.), *Vocabulary acquisition: Implications for reading comprehension* (pp. 233–248). NY: The Guilford Press.

Gathercole, S., & Baddeley, A. (1993). Phonological working memory: A critical building block for reading development and vocabulary acquisition? *European Journal of Pscyhology of Education, 8*(3), 259–272.
Gentry, J. R. (2000). A retrospective on invented spelling and a look forward. *The Reading Teacher, 54*(3), 318–332.
Graham, S. (2000). Should the natural learning approach replace spelling instruction? *Journal of Educational Psychology, 92*(2), 235–247. https://doi.org/10.1037//0022-0663.92.2.235.
Graham, S., Harris, K., & Adkins, M. (2018). The impact of supplemental handwriting and spelling instruction with first grade students who do not acquire transcription skills as rapidly as peers: a randomized control trial. *Reading and Writing, 31*(6), 1273–1294. https://doi.org/10.1007/s11145-018-9822-0.
Graham, S., & Santangelo, T. (2014). Does spelling instruction make students better spellers, readers, and writers? *A meta-analytic review. Reading and Writing, 27*(9), 1703–1743. https://doi.org/10.1007/s11145-014-9517-0.
Green, L., McCutchen, D., Schwiebert, C., Quinlan, T., Eva-Wood, A., & Juelis, J. (2003). Morphological development in children's writing. *Journal of Educational Psychology, 95*(4), 752–761.
Hambleton, R. K., & Jones, R. W. (1993). *Comparison of classical test theory and item response theory and their applications to test development* (pp. 39–47). Educational Measurement: Issues and Practice.
Kohnen, S., Colenbrander, D., Krajenbrink, T., & Nickels, L. (2015). Assessment of lexical and nonlexical spelling in students in Grades 1–7. *Australian Journal of Learning Difficulties, 20*(1), 15–38. https://doi.org/10.1080/19404158.2015.1023209.
Kohnen, S., Nickels, L., & Castles, A. (2009). Assessing spelling skills and strategies: A critique of available resources. *Australian Journal of Learning Difficulties, 14*(1), 113–150.
Kohnen, S., Nickels, L., Castles, A., Friedmann, N., & McArthur, G. (2012). When 'slime' becomes 'smile': Developmental letter position dyslexia in English. *Neuropsychologia, 50*(14), 3681–3692. https://doi.org/10.1016/j.neuropsychologia.2012.07.016.
Kottner, J., Audig, L., Brorson, S., Donner, A., Gajewski, B. J., Hrobjartsson, A., et al. (2011). Guidelines for reporting reliability and agreement studies (GRRAS) were proposed. *Journal of Clinical Epidemiology, 64,* 96–106.
Larkin, R., Williams, G., & Blaggan, S. (2013). Delay or deficit? Spelling processes in children with specific language impairment. *Journal of Communication Disorders, 46,* 401–412. https://doi.org/10.1016/j.jcomdis.2013.07.003.
Linacre, J. M. (2002). What do infit and outfit, mean-square and standardized mean? *Rasch Measurement Transactions, 16*(2), 878.
Martin-Chang, S., Ouellette, G., & Madden, M. (2014). Does poor spelling equate to slow reading? The relationship between reading, spelling, and orthographic quality. *Reading and Writing, 27*(8), 1485–1505. https://doi.org/10.1007/s11145-014-9502-7.
Masso, S., & Baker, E. (2015). Phonology. In J. McLeod & J. McCormack (Eds.), *Introduction to speech, language and literacy* (pp. 134–177). South Melbourne: Oxford University Press.
McLeod, J., & McCormack, J. (Eds.). (2015). *Introduction to speech, language and literacy*. South Melbourne: Oxford University Press.
Moats, L. (2005/06). How spelling supports reading. *American Educator, Winter*, *6*, 12–43.
Muijs, D. (2004). Validity, reliability and generalisability. In *Doing quantitative research in education with SPSS* (pp. 64–76). London: Sage Publications.
Nunes, T., & Bryant, P. (2006). *Improving literacy by teaching morphemes*. London: Routledge.
Nunes, T., Bryant, P., & Olsson, J. (2003). Learning morphological and phonological spelling rules: An intervention study. *Scientific Studies of Reading, 7*(3), 289–307. https://doi.org/10.1207/s1532799xssr0703_6.
Plaza, M., & Cohen, H. (2003). The interaction between phonological processing, syntactic awareness, and naming speed in the reading and spelling performance of first-grade children. *Brain and Cognition, 53*(2), 287–292. https://doi.org/10.1016/S0278-2626(03)00128-3.
Richards, T., Aylward, E., Field, K., Grimme, A., Raskind, W., Richards, A., et al. (2006). Converging evidence for Triple Word Form Theory in children with dyslexia. *Developmental Neuropsychology, 30*(1), 547–589. https://doi.org/10.1207/s15326942dn3001_3.
Rittle-Johnson, B., & Siegler, R. S. (1999). Learning to spell: Variability, choice, and change in children's strategy use. *Child Development, 70*(2), 332–348.

Ruberto, N., Daigle, D., & Ammar, A. (2016). The spelling strategies of francophone dyslexic students. *Reading and Writing, 29*(4), 659–681. https://doi.org/10.1007/s11145-015-9620-x.
Sacre, L., & Masterson, J. (2000). *Single word spelling test*. London: Nfer Nelson.
Siegel, L. (2008). Phonological processing deficits and reading disabilities. In J. Metsala & L. Ehri (Eds.), *Word recognition in beginning literacy* (pp. 141–160). Mahwah, NJ: Lawrence Erlbaum.
Sireci, S., & Falkner-Bond, M. (2014). Validity evidence based on test content. *Psicothema, 26*(1), 100–107.
Snowling, M., & Hulme, C. (2012). Annual research review: The nature and classification of reading disorders—A commentary on proposals for DSM-5. *The Journal of Child Psychology and Psychiatry, 53*(5), 593–607. https://doi.org/10.1111/j.1469-7610.2011.02495.x.
Sumner, E., Connelly, V., & Barnett, A. (2016). The influence of spelling ability on vocabulary choices when writing for children with dyslexia. *Journal of Learning Disabilities, 49*(3), 293–304.
Treiman, R. (2017a). Learning to spell words: Findings, theories, and issues. *Scientific Studies of Reading, 21*(4), 1–12. https://doi.org/10.1080/10888438.2017.1296449.
Treiman, R. (2017b). Learning to spell: Phonology and beyond. *Cognitive Neuropsychology*. https://doi.org/10.1080/02643294.2017.1337630.
Treiman, R. (2018). Teaching and learning spelling. *Child Development Perspectives, 12*(3), 1–5. https://doi.org/10.1111/cdep.12292.
Treiman, R., & Kessler, B. (2006). Spelling as statistical learning: Using consonantal context to spell vowels. *Journal of Educational Psychology, 98*(3), 642–652.
Varnhagen, C., McCallum, M., & Burstow, M. (1997). Is children's spelling naturally stage-like? *Reading and Writing, 9,* 451–481. https://doi.org/10.1023/A:1007903330463.
Wagner, R. K., Torgesen, J. K., Rashotte, C. A., & Pearson, N. A. (2013). *Comprehensive test of phonological processing* (2nd ed.). Austin, TX: Pro Ed.
Westwood, P. (2005). *Spelling: Approaches to teaching and assessment*. Camberwell, VIC: ACER Press.