



UNIVERSITY  
OF WOLLONGONG  
AUSTRALIA

2014

# Assessing what students know: Effects of assessment type on spelling performance and relation to working memory

Alysha M. Calleia

University of Wollongong, amc998@uowmail.edu.au

Steven J. Howard

University of Wollongong, stevenh@uow.edu.au

Follow this and additional works at: <http://ro.uow.edu.au/jseem>



Part of the [Education Commons](#)

### Recommended Citation

Calleia, Alysha M. and Howard, Steven J., Assessing what students know: Effects of assessment type on spelling performance and relation to working memory, *Journal of Student Engagement: Education Matters*, 4(1), 2014, 14-24.

Available at: <http://ro.uow.edu.au/jseem/vol4/iss1/3>

---

# Assessing what students know: Effects of assessment type on spelling performance and relation to working memory

## **Abstract**

A central objective of educational assessment is to maximise the accuracy (validity) and consistency (reliability) of the methods used to assess students' competencies. Different tests, however, often employ different methods of assessing the same domain-specific skills (e.g., spelling). As a result, questions have arisen concerning the legitimacy of using these various modes interchangeably as a proxy for students' abilities. To investigate the merit of these contentions, this study examined university students' spelling performance across three commonly employed test modalities (i.e., dictation, error correction, proofreading). To further examine whether these test types vary in the cognitive load they place on test takers, correlations between working memory and spelling scores were also examined. Results indicated that the modes of assessment were not equivalent indices of individuals' orthographic knowledge. Specifically, performance in the dictation and error correction conditions were superior to that in the proofreading condition. Moreover, correlational analyses revealed that working memory accounted for significant variance in performance in the dictation and error correction conditions (but not in the proofreading condition). These findings suggest that not all standardised assessment methods accurately capture student competencies and that these domain-specific assessments should seek to minimise the domain-general cognitive demands placed on test takers.

## **Keywords**

spelling, assessment, working memory, cognitive load, validity, reliability



## **Assessing what students know: Effects of assessment type on spelling performance and relation to working memory**

Alysha M. Calleia

*Bachelor of Primary Education (fourth year), School of Education, University of Wollongong, Australia*

Dr Steven J. Howard\*

*School of Education, University of Wollongong, Australia*

(\*Originator and supervisor of lead author's undergraduate research study)

A central objective of educational assessment is to maximise the accuracy (validity) and consistency (reliability) of the methods used to assess students' competencies. Different tests, however, often employ different methods of assessing the same domain-specific skills (e.g., spelling). As a result, questions have arisen concerning the legitimacy of using these various modes interchangeably as a proxy for students' abilities. To investigate the merit of these contentions, this study examined university students' spelling performance across three commonly employed test modalities (i.e., dictation, error correction, proofreading). To further examine whether these test types vary in the cognitive load they place on test takers, correlations between working memory and spelling scores were also examined. Results indicated that the modes of assessment were not equivalent indices of individuals' orthographic knowledge. Specifically, performance in the dictation and error correction conditions were superior to that in the proofreading condition. Moreover, correlational analyses revealed that working memory accounted for significant variance in performance in the dictation and error correction conditions (but not in the proofreading condition). These findings suggest that not all standardised assessment methods accurately capture student competencies and that these domain-specific assessments should seek to minimise the domain-general cognitive demands placed on test takers.

**Keywords:** spelling, assessment, working memory, cognitive load, validity, reliability

### **Introduction**

Maximising accuracy (validity) and consistency (reliability) of the methods used to measure students' achievements, expertise and abilities remain central goals of educational assessment (Pellegrino, 2001). As a result, the past two decades have been characterised by a rapid increase in the development, rigorous psychometric evaluation and implementation of standardised tests, many of which involve high-stake consequences for educators and schools (Chudowsky & Pellegrino, 2003; Pellegrino, 2006). In Australia, for instance, schools are ranked based on their pupils'



overall performance on tests such as the National Assessment Program – Literacy and Numeracy (NAPLAN; Masters, 2013). Internationally, teachers' pay and job security are also often determined by their students' performance on national or state-mandated standardised tests (Morris, 2011; Wang, Beckett & Brown, 2006). The high stakes associated with testing emphasises the need for assessments that accurately index students' competencies.

Despite this trend toward standardisation, these tests typically employ different methods of assessing the same domain-specific skills. For instance, standardised spelling assessments often employ one of four modes of assessment to index learners' levels of orthographic knowledge (i.e., proofreading, error correction, multiple choice and/or dictation; Moats, 2005; Pearson, 2012). Given evidence that students' test performance differs across these modalities (Croft, 1982; Frisbie & Cantor, 1995; Willet & Gardiner, 2009), questions have arisen concerning the legitimacy of using these various modes interchangeably as a proxy for students' abilities (Pellegrino, 2006; Wiliam, 2003). That is, if common standardised tests are equally accurate and consistent measures of students' spelling competencies, performance on these tests should be highly similar. However, Willet and Gardiner's (2009) testing of 2,369 primary school students revealed that over 75% of participants spelt more words correctly when dictated compared to their spelling performance on NAPLAN (which involves error correction and proofreading). The researchers explained this result by suggesting that incorrect letter sequences may have interfered with students' abilities to produce the correct spelling of the target word (e.g., incorporating the spelling error into their own response). These findings are also consistent with earlier studies, which suggest that error correction and proofreading tests may require skills over and above spelling ability (Croft, 1982; Frisbie & Cantor, 1995).

This possibility remains debated, however, in light of research indicating that assessment type has little to no effect on examinees' performance (e.g., Allred, 1984; Andrews & Scarratt, 1996; Moseley, 1997; Westwood, 1999). Westwood (1999), for example, reported high correlations between 93 second to fifth-grade students' spelling scores on proofreading, dictation, multiple-choice, error correction and authentic writing tasks (correlations between test types ranged from .77 to .97). Although such evidence suggests that these tests are relatively equivalent indices of students' proficiencies, it is noted that the proofreading task employed in this study required students to only identify the misspelt word rather than correct it (which may not accurately index the students' spelling competencies). Accordingly, different task requirements across studies may explain these discrepant findings.

### **The cognitive load of assessment**

One possible explanation for variable performance across different spelling conditions is that different modes of assessment may vary in the cognitive load they place on the test-taker's working memory. Working memory refers to our capacity-limited mental workspace for the temporary maintenance and manipulation of information (Alloway, 2006; Engle, 2010). Given that only a limited amount of information can concurrently be activated in working memory, tasks that require individuals to hold and manipulate larger amounts of information are more cognitively demanding and, therefore, more



difficult than tasks involving less mental effort (Ricker, AuBuchon & Cowan, 2010). Extending this principle to spelling, Pearson (2012) postulated that because proofreading and error correction tasks require individuals to shift from one orthographic representation to another, this two-step process may place greater demands on students' working memory than tasks requiring students to recall the correct letter sequence with little interference from external factors. Similarly, as dictation tests involve fewer component processes (e.g., there is no need to proofread or correct), it can be postulated that relatively less mental attention would be needed to perform such tasks.

Although there is no explicit empirical evidence to date supporting these assertions, there is extensive evidence that working memory underlies performance on numerous standardised assessments (e.g., Ackerman, Beier & Boyle, 2005; Alloway & Gregory, 2012; Best, Miller & Naglieri, 2011; Gathercole et al., 2003; Lan et al., 2011). For example, a study examining the relationship between primary school students' working memory skills and their performance on English, Mathematics and Science Scholastic Aptitude Tests (SATs) demonstrated that working memory was moderately to highly correlated with performance on all tests (correlation ranged between .36 to .53; Gathercole et al., 2003). In addition, working memory is significantly related to students' spelling abilities (as measured by performance on standardised spelling subtests; Alloway & Gregory, 2012; Strattman & Hodson, 2005). In fact, Alloway and Gregory (2012) found working memory to be the strongest predictor of spelling performance. Taken together, these results suggest that the cognitive load of different assessment types may unduly influence students' performance on these tests, thus obscuring the accurate measurement of students' competencies and capacities.

### **The importance of accurate spelling assessment**

Although spelling has been considered a somewhat 'constrained skill' by some (e.g., Paris, 2005), the importance of spelling has been emphasised by research indicating that proficient orthographic knowledge provides the foundation for the reading and writing skills necessary to become a literate individual (Ehri, 2000; Kelman & Apel, 2004; Masterson & McLaughlin, 2009). Further, students' spelling abilities have been found to be highly related to their reading accuracy, vocabulary knowledge, written expression and comprehension skills (August, 2011; Moats, 2005; Nagy & Scott, 2000; Puranik & Al Otaiba, 2012). Likewise, deficits in spelling skills negatively impact the quality and length of students' written compositions (Aram, 2005; Graham et al., 1997; Graham, Harris & Chorzempa, 2002), as well as their ability to fluently decode texts (Berninger et al., 1998; Gentry, 2004). Accurate and reliable measures of students' spelling competencies are, therefore, imperative to help educators determine instructional needs to support students' reading and writing achievement.

Given the vital role of spelling in fostering literate individuals, yet the lack of clarity that surrounds the comparative reliability and validity of different spelling assessments, additional research is needed to determine whether different modes of assessment provide equally accurate measurements of students' orthographic knowledge. The present study thus sought to investigate the effects of assessment type on spelling performance by examining differences in participants' scores across three



common test modalities (i.e., dictation, error correction, proofreading). These spelling modalities were chosen for consistency with Australia's NAPLAN testing (i.e., error correction, proofreading), the UK's Grammar and Phonology Screening test (i.e., dictation) and many commercially available standardised spelling assessments (e.g., Boder Test of Reading, Wide Range Achievement Test 4 (WRAT-4), Wechsler Individual Achievement Test Second Edition (WIAT-II), South Australian Spelling Test). Further, to determine whether these differences are related to the cognitive load of each assessment, the relationship between participants' working memory capacity (as indexed by participants' performance on the Figural Intersection Task) and their performance in each spelling condition was examined. In line with previous research, it was expected that the dictation condition would yield the highest level of performance. In addition, based on the presumption that increasing the sophistication and complexity of the task increases the demands placed on test-takers' working memory (Pearson, 2012), it was expected that working memory would correlate with all tasks, yet more highly so with the proofreading condition.

## **Method**

### ***Participants***

Participants were 31 third-year undergraduate Bachelor of Primary Education students. All students were enrolled in either an educational literacy or research subject at the same New South Wales university and volunteered in response to the circulation of participant information sheets. This sample consisted of 6 males and 25 females whose age ranged from 20 to 39 years ( $M = 24.13$ ;  $SD = 5.03$ ). All participants were native speakers of English.

### ***Measures***

#### ***Spelling tasks***

In order to investigate the effects of assessment type on spelling performance, three spelling tests that differed in item format were administered. For the *proofreading* test, participants identified and corrected the misspelt word in each sentence (e.g., "A solar eclipse is a rare ocurrence"). For the *error correction* test, participants corrected the identified misspelt word in each sentence (e.g., "A solar eclipse is a rare ocurrence"). For the *dictation* test, participants spelt the missing word in each sentence (e.g., "A solar eclipse is a rare \_\_\_\_\_"). In all cases, sentences were presented both visually and auditorily on a laptop computer using E-Prime 2.0 software (Psychology Software Tools, Pittsburgh, PA).

Each spelling test consisted of ten words randomly selected from a common pool of 90 words. The 90 stimulus words were selected from standardised tests with spelling subscales (i.e., WIAT-II, WRAT-4, Boder Test of Reading), although the contextual sentences were modified from their original form. Words were selected to be age-appropriate for adults, yet provide a range of item difficulties (to prevent frustration and enhance motivation). Word stimuli were adapted for the proofreading and error correction conditions by misspelling each word in one of the following ways: omitting one or more letters (23.3%); adding one or more letters (5.6%);



substituting single letters, digraphs or diphthongs (37.8%); presenting the incorrect homophone (3.3%); writing the incorrect letter order (5.6%); using phonetic spelling inappropriately (4.4%); not abiding by the rule/convention (4.4%) or using a combination of these (15.6%). These misspellings were based on those identified by Roberts (2003) as the most common error types. Words were never repeated within a testing session; as such, each participant was never presented with the same word twice.

Prior to commencing each spelling test, participants were provided with instructions that identified the upcoming spelling condition and what this required of them. Participants were then presented with ten sentences of the same type (i.e., proofreading, error correction or dictation) and responded to each sentence by writing their spelling of the target word on a scoring sheet. Spelling performance was indexed by accuracy scores for each test condition. Items were scored dichotomously as correct (1) or incorrect (0) spellings. In the proofreading condition, participants received a score of 1 only if they identified the misspelt word *and* spelt it correctly.

#### *Figural Intersection Task (FIT)*

An iPad-based version of the FIT (Pascual-Leone & Baillargeon, 1994; Pascual-Leone & Johnson, 2011) was used to measure participants' working memory capacity. This task consists of eight training items and 42 test items, each of which present an array of two to eight figures on the right-hand side (the relevant shapes) and an overlapping configuration of the same shapes (and, in some instances, additional irrelevant shapes to be ignored) on the left. Using a stylus, participants were required to place a single dot in the area of common intersection of the relevant overlapping shapes on the left. The difficulty level (or item class) of each item was defined by the number of relevant shapes to be intersected. There were six items in each item class. Working memory capacity was indexed by the highest item class for which participants could accurately complete at least 80% of the items (as long as this 80% threshold was also achieved for all lower classes, with one item class permitted to fall to a 60% threshold; Howard, Johnson & Pascual-Leone, 2013).

#### ***Procedure***

All measures were administered individually, in a single testing session (~60 minutes). This testing took place in a quiet office at the university. Spelling tests were administered first, in a counterbalanced order, followed by administration of the FIT.

#### ***Results***

A Greenhouse-Geisser repeated measures ANOVA was conducted on the number of words correctly spelt within each test condition. Exploration of the data revealed three extreme observations, which were ultimately retained because their exclusion did not affect the overall pattern of results. Contrary to expectations, results indicated no significant effect of test type,  $F(1.94, 58.14) = 2.31$ ,  $p = .110$ , partial  $\eta^2 = .07$ , suggesting that the mode of assessment had no significant effect on spelling performance. Examination of descriptive statistics, however, suggested that participants performed better (although non-significantly so) in the dictation condition



( $M = 8.06$ ,  $SD = 1.71$ ) and error correction condition ( $M = 8.00$ ,  $SD = 1.48$ ), compared to the proofreading condition ( $M = 7.39$ ,  $SD = 1.81$ ). This suggests that the lack of statistical significance may be a function of insufficient statistical power (i.e., small sample size) to detect a genuine difference in performance.

To assess the potentially variable relationship between working memory and performance on the different spelling tests, correlations between spelling scores and FIT scores were computed. Results revealed significant positive correlations for the FIT with performance in the error correction condition,  $r(28) = .44$ ,  $p = .017$ ,  $R^2 = .19$ , and dictation condition,  $r(28) = .46$ ,  $p = .011$ ,  $R^2 = .21$ . This suggests that better working memory performance was associated with better spelling performance in these conditions (and vice versa). Contrary to expectations, however, there was no significant correlation between FIT and performance in the proofreading condition,  $r(28) = .22$ ,  $p = .262$ ,  $R^2 = .05$ . Nevertheless, the presence of significant correlations provides at least tentative evidence that different test types place differing demands on individuals' working memory.

To further evaluate this possibility, correlations between spelling conditions were also computed. That is, if all test types were equally valid and reliable measures of students' spelling abilities, a largely similar performance (a high degree of correlation) could be expected across test types. However, results indicated that all three test types were only moderately correlated: error correction–proofreading,  $r(30) = .40$ ,  $p = .028$ ,  $R^2 = .16$ ; error correction–dictation,  $r(30) = .33$ ,  $p = .071$ ,  $R^2 = .11$ ; proofreading–dictation,  $r(30) = .29$ ,  $p = .112$ ,  $R^2 = .08$ . Results thus suggest that: (i) these tests did not appear to be measuring the same underlying competency (that is, the ability to spell may not be synonymous with the ability to correct misspellings); and/or (ii) these tests did not appear to be measuring spelling in the same way (e.g., they differed in the additional skills – those extraneous to ability to spell a word – required to successfully spell the target word).

## **Discussion**

The current study aimed to investigate whether spelling performance was influenced by the mode of assessment (i.e., dictation, error correction, proofreading) and whether performance on these tests was associated with the cognitive load placed on test-takers' working memory. That is, if all standardised tests are equally valid and reliable measures of students' spelling competencies, performance on these tests should be highly similar and correlations between test scores should be large. However, the current results provide evidence that the three assessment types varied in their estimates of participants' orthographic knowledge (as highlighted by non-significant differences in accuracy scores). Further, there were only moderate correlations between test scores. The variable relationship between working memory and spelling performance (which depended on the mode of assessment) suggests that this variability in spelling performance may have been a function of the differing cognitive load imposed on test takers.

Taken together, these results highlight the problematic nature of interpreting performance on standardised assessments as equivalent indices of students' orthographic knowledge. Specifically, in the current study, proofreading scores were deflated relative to the competencies displayed by participants in the dictation and



error correction conditions. This finding is directionally consistent with research indicating that dictation tasks typically yield superior spelling performance (Croft, 1982; Frisbie & Cantor, 1995; Willet & Gardiner, 2009). Given that these tasks require students to spell the target word, this would suggest that dictation-type spelling tests might provide more-accurate estimates of students' underlying spelling competencies. The current results supplement this by suggesting that simple error correction may provide an equally accurate means of spelling assessment.

That this variability in performance is related to the cognitive load placed on test takers was supported by significant correlations between working memory performance and spelling scores. Although there is little prior research in this area, this finding is consistent with research showing working memory as an important factor underlying achievement on various standardised tests (e.g., Ackerman, Beier & Boyle, 2005; Alloway & Gregory, 2012; Gathercole et al., 2003; Lan et al., 2011; Strattman & Hodson, 2005). In the current study, working memory accounted for significant variance in performance on both the dictation and error correction conditions. Contrary to expectations, however, only a small non-significant correlation was found between working memory and proofreading scores. Due to the similarities between this condition and the error correction condition (the only difference being the need to identify the misspelt word), as well as scores being lowest on the proofreading test (suggesting that it was indeed more difficult), it may be that this result is spurious and/or a product of limitations in the study design (e.g., sample size). Further research is required to investigate these possibilities.

Nevertheless, the moderate correlations between spelling scores in the current study support the assertion that these test types vary in their complexity. This is problematic given the assertion that for assessments to be considered valid “the underlying unobservable hypothetical construct should correlate with performance in other tasks for which that construct is deemed important” (Engle, 2010, p. S20). By extension, any standardised method of assessing spelling should strongly predict performance on any other standardised test of spelling. However, that spelling scores were not highly correlated in the current study implies that these assessments were not measuring the same underlying competency. Although this finding contradicts previous research findings that different forms of spelling assessment yield highly correlated scores (e.g., Allred, 1984; Andrews & Scarratt, 1996; Moseley, 1997; Westwood, 1999), this discrepancy is likely attributable to differences in test format. For instance, whereas some studies have only required participants to identify misspelt words to index their proofreading ability (e.g., Westwood, 1999), the current study additionally required participants to correct the misspelt word (in line with NAPLAN's procedures). This added component likely resulted in the proofreading task being more challenging in the current study (Pearson, 2012). This difference in cognitive load imposed by the differing test types may have contributed to some of the variability in spelling performance (although the small correlation with the proofreading condition remains curious in this regard).

These results should be interpreted in the context of a number of limitations. For one, the current study is limited by its small sample size, in which even slight fluctuations in the data can influence the pattern of results (e.g., increasing the chance of a type II error; Field, 2013). Moreover, randomisation of words in each spelling test left the difficulty of each test to chance. As a result, some participants may have



completed tests that were substantially easier or harder relative to the other conditions. In fact, an investigation of list difficulty (identified by standardised test classifications and overall accuracy scores for each item) indicated that in many instances at least one test condition was allocated significantly easier or harder words compared to the other two conditions. Finally, as the study employed a version of the FIT that is still in the validation stages, it is unclear whether this task provided an accurate index of participants' working memory capacities. Further research to address these methodological shortcomings is required.

## **Conclusion**

Despite the limitations, the current findings raise important questions concerning the legitimacy of interpreting common standardised assessments as interchangeable indices of student competencies. Specifically, the current study provides evidence that these modes of spelling assessment might not be equally valid and reliable measures of individuals' orthographic knowledge. Rather, the current results suggest that dictation tasks may yield better estimates of students' emerging competencies. In addition, the current results provide evidence that at least some of the variability in spelling scores may be related to the cognitive load placed on the test taker. That is, cognitive load theorists currently acknowledge that different instructional designs vary in the cognitive demands they place on learners' working memory (e.g., Paas, Renkl & Sweller, 2004; Paas et al., 2003). The current study furthers this idea by suggesting that educational assessments also vary in the cognitive demands they place on test takers. This suggests that assessment design may also benefit from cognitive load-compliant design. Specifically, assessments should be designed to manage intrinsic load (the inherent complexity of the task), minimise extraneous load (information and processes that are non-essential for learning) and maximise germane load (learning-relevant processes; Sweller, 2010). Although further research is required to replicate these results, and extend them to additional domains of assessment (e.g., numeracy, reading), current findings emphasise the need for assessments that optimise the load placed on test-takers' working memory. Essentially, taking a theory-guided approach to assessment design may provide a means to maximise the accuracy and reliability not only of spelling assessments, but also of educational assessments more broadly.

## **References**

- Ackerman, P.L., Beier, M.E. & Boyle, M.O. (2005). 'Working memory and intelligence: The same or different constructs?' *Psychological Bulletin*, 131 (1), 30–60.
- Alloway, T.P. (2006). 'How does working memory work in the classroom?' *Educational Research and Reviews*, 1 (4), 134–139.
- Alloway, T.P. & Gregory, D. (2012). 'The predictive ability of IQ and working memory scores in literacy in an adult population'. *International Journal of Educational Research*, 57 (1), 51–56.



- Allred, R.A. (1984). 'Comparison of proofreading-type standardised spelling tests and written spelling test scores'. *Journal of Educational Research*, 77 (5), 298–303.
- Andrews, S. & Scarratt, D.R. (1996). 'What comes after phonological awareness? Using lexical experts to investigate orthographic processes in reading'. *Australian Journal of Psychology*, 48 (3), 141–148.
- Aram, D.M. (2005). 'Continuity in children's literacy achievements: A longitudinal perspective from kindergarten to school'. *First Language*, 25 (3), 259–289.
- August, G. (2011). 'Spelling facilitates good ESL reading comprehension'. *Journal of Developmental Education*, 35 (1), 14–16.
- Berninger, V.W., Vaughn, K., Abbott, R., Brooks, A., Rogan, L. & Reed, E. (1998). 'Early intervention for spelling problems: Teaching functional spelling units of varying size with a multiple-connections framework'. *Journal of Educational Psychology*, 90 (4), 587–605.
- Best, J.R., Miller, P.H. & Naglieri, J.A. (2011). 'Relations between executive function and academic achievement from ages 5 to 17 in a large, representative sample'. *Learning Individual Differences*, 21 (4), 327–336.
- Boder, E. & Jarrico, S. (1982). *The Boder Test of Reading – Spelling Patterns: A Diagnostic Screening Test for Subtypes of Reading Disability*. New York: Grune & Stratton.
- Chudowsky, N. & Pellegrino, J.W. (2003). 'Large-scale assessments that support learning: What will it take?' *Theory into Practice*, 42 (1), 75–83.
- Croft, A.C. (1982). 'Do spelling tests measure the ability to spell?' *Educational and Psychological Measurement*, 42 (3), 715–723.
- Ehri, L.C. (2000). 'Learning to read and learning to spell: Two sides of a coin'. *Topics in Language Disorders*, 20 (3), 19–36.
- Engle, R.W. (2010). 'Role of working-memory capacity in cognitive control'. *Current Anthropology*, 51 (S1), S17–S26.
- Field, A. (2013). *Discovering Statistics using IBM SPSS Statistics and Sex and Drugs and Rock 'n' Roll*, 4th edn. London: SAGE.
- Frisbie, D.A. & Cantor, N.K. (1995). 'The validity of scores from alternative methods of assessing spelling achievement'. *Journal of Educational Measurement*, 32 (1), 55–78.
- Gathercole, S.E., Pickering, S.J., Knight, C. & Stegmann, Z. (2003). 'Working memory skills and educational attainment: Evidence from national curriculum assessments at 7 and 14 years of age'. *Applied Cognitive Psychology*, 18 (1), 1–16.
- Gentry, J.R. (2004). *The Science of Spelling: The explicit specifics that make great readers and writers (and spellers!)*. Portsmouth, NH: Heinemann.
- Graham, S., Berninger, V., Abbott, R., Abbott, S. & Whitaker, D. (1997). 'Role of mechanics in composing of elementary school students: A new methodological approach'. *Journal of Educational Psychology*, 89 (1), 170–182.
- Graham, S., Harris, K. & Chorzempa, B.F. (2002). 'Contribution of spelling instruction to the spelling, writing, and reading of poor spellers'. *Journal of Educational Psychology*, 94 (4), 669–686.
- Howard, S.J., Johnson, J. & Pascual-Leone, J. (2013). 'Measurement of mental attention: Assessing a cognitive component underlying performance on



- standardized intelligence tests'. *Psychological Test and Assessment Modeling*, 55 (3), 250–273.
- Kelman, M. & Apel, K. (2004). 'The effects of a multiple linguistic, prescriptive approach to spelling instruction: A case study'. *Communication Disorders Quarterly*, 25 (2), 56–66.
- Lan, X., Legare, C.H., Ponitz, C.C., Li, S. & Morrison, F.J. (2011). 'Investigating the links between the subcomponents of executive function and academic achievement: A cross-cultural analysis of Chinese and American pre-schoolers'. *Journal of Experimental Child Psychology*, 108 (3), 677–692.
- Masters, G.N. (2013). *Reforming Educational Assessment: Imperatives, principles and challenges*. Camberwell, VIC: ACER Press.
- Masterson, J. & McLaughlin, V. (2009). The relationship between spelling and reading: The metric matters. Poster presented at the annual meeting of the Society for the Scientific Study of Reading, 25<sup>th</sup> June, Boston, MA.
- Moats, L.C. (2005). 'How spelling supports reading: And why it is more regular and predictable than you may think'. *American Educator*, 29 (4), 12–22.
- Morris, A. (2011). *Student Standardised Testing: Current practices in OECD countries and a literature review*. OECD Education Working Papers, No. 65. Paris: OECD. URL: [www.oecd.org/edu/workingpapers](http://www.oecd.org/edu/workingpapers) (accessed 07 August 2013).
- Moseley, D.V. (1997). 'Assessment of spelling and related aspects of written expression'. In J.R. Beech & C. Singleton (eds), *The Psychological Assessment of Reading* (pp. 204–223). London: Routledge.
- Nagy, W.E. & Scott, J.A. (2000). 'Vocabulary processes'. In M.L. Kamil, P.B. Mosenthal, P.D. Pearson & R. Barr (eds), *Handbook of Reading Research*, vol. 3, (pp. 269–284). Mahwah, NJ: Lawrence Erlbaum.
- Paas, F., Renkl, A. & Sweller, J. (2004). 'Cognitive Load Theory: Instructional implications of interaction between information structures and cognitive architecture'. *Instructional Science*, 32 (1), 1–8.
- Paas, F., Tuovinen, J., Tabbers, H. & Van Gerven, P. (2003). 'Cognitive load measurement as a means to advance cognitive load theory'. *Educational Psychologist*, 38 (1), 63–71.
- Paris, S.G. (2005). 'Reinterpreting the development of reading skills'. *Reading Research Quarterly*, 40 (2), 184–202.
- Pascual-Leone, J. & Baillargeon, R. (1994). 'Developmental measurement of mental attention'. *International Journal of Behavioral Development*, 17 (1), 161–200.
- Pascual-Leone, J. & Johnson, J. (2011). 'A developmental theory of mental attention'. In P. Barrouillet & V. Gaillard (eds), *Cognitive Development and Working Memory: From neo-Piagetian to cognitive approaches* (pp. 13–46). New York: Psychology Press.
- Pearson, H. (2012). 'Issues in the assessment of spelling'. *Literacy Learning: The Middle Years*, 20 (2), 29–33.
- Pellegrino, J.W. (2006). *Rethinking and Redesigning Curriculum, Instruction and Assessment: What contemporary research and theory suggests*. Washington, DC: National Center on Education and the Economy. URL: <http://www.skillscommission.org/wp-content/uploads/2010/05/Rethinking-and-Redesigning.pdf> (accessed 02 August 2013).



- Puranik, C. & Al Otaiba, S. (2012). 'Examining the contribution of handwriting and spelling to written expression in kindergarten children'. *Reading and Writing: An Interdisciplinary Journal*, 25, (7) 1523–1546.
- Ricker, T. J., AuBuchon, A. & Cowan, N. (2010). Working memory. *Wiley Interdisciplinary Reviews: Cognitive Science*, 1 (4), 573–585.
- Roberts, J. (2001). 'Errors, causes and actions to take'. In J. Roberts (ed.), *Spelling Recovery: The pathway to spelling success* (pp. 1–13). Melbourne, VIC: ACER Press.
- Strattman, K. & Hodson, B.W. (2005). 'Variables that influence decoding and spelling in beginning readers'. *Child Language Teaching and Therapy*, 21 (2), 165–190.
- Sweller, J. (2010). 'Element interactivity and intrinsic, extraneous and germane cognitive load'. *Educational Psychological Review*, 22 (2), 123–138.
- Wang, L., Beckett, G. & Brown, L. (2006). 'Controversies of standardized assessment in school accountability reform: A critical synthesis of multidisciplinary research evidence'. *Applied Measurement in Education*, 19 (4), 305–328.
- Wechsler, D. (2005). *Wechsler Individual Achievement Test 2nd Edition (WIAT II)*. London: Psychological Corp.
- Westwood, P. (1999). 'The correlation between results from different types of spelling test and children's spelling ability when writing'. *Australian Journal of Learning Disabilities*, 4 (1), 31–36.
- Wiliam, D. (2003). 'National curriculum assessment: How to make it better'. *Research Papers in Education*, 18 (2), 129–136.
- Wilkinson, G.S. & Robertson, G.J. (2006). *Wide Range Achievement Test–Fourth Edition (WRAT-4)*. Lutz, FL: Psychological Assessment Resources.
- Willet, L. & Gardiner, A. (2009). Testing spelling: Exploring NAPLAN. Paper presented at Australian Literacy Educators Association Conference, Hobart, 9–12 July. URL: <http://www.englishliteracyconference.com.au/files/documents/hobart/conferencePapers/nonRefereed/Willet-Gardiner.pdf> (accessed 05 August 2013).