*Neuroscience* 297 (2015) 38–46

# BEHAVIORAL AND fMRI EVIDENCE OF THE DIFFERING COGNITIVE LOAD OF DOMAIN-SPECIFIC ASSESSMENTS

**S. J. HOWARD,[a*] H. BURIANOVÁ,[b] J. EHRICH,[c] L. KERVIN,[a] A. CALLEIA,[a] E. BARKUS,[d] J. CARMODY[e] AND S. HUMPHRY[f]**

[a] *School of Education, University of Wollongong, New South Wales 2522, Australia*

[b] *Centre for Advanced Imaging, University of Queensland, Queensland 4072, Australia*

[c] *Faculty of Education, Monash University, Victoria 3800, Australia*

[d] *School of Psychology, University of Wollongong, New South Wales 2522, Australia*

[e] *Neurology Department, Wollongong Hospital, New South Wales 2500, Australia*

[f] *Graduate School of Education, University of Western Australia, Western Australia 6009, Australia*

**Abstract—Standards-referenced educational reform has increased the prevalence of standardized testing; however, whether these tests accurately measure students' competencies has been questioned. This may be due to domain-specific assessments placing a differing domain-general cognitive load on test-takers. To investigate this possibility, functional magnetic resonance imaging (fMRI) was used to identify and quantify the neural correlates of performance on current, international standardized methods of spelling assessment. Out-of-scanner testing was used to further examine differences in assessment results. Results provide converging evidence that: (a) the spelling assessments differed in the cognitive load placed on test-takers; (b) performance decreased with increasing cognitive load of the assessment; and (c) brain regions associated with working memory were more highly activated during performance of assessments that were higher in cognitive load. These findings suggest that assessment design should optimize the cognitive load placed on test-takers, to ensure students' results are an accurate reflection of their true levels of competency.** 

**Key words: cognitive load, standardized testing, assessment, spelling, fMRI, educational neuroscience.**

## INTRODUCTION

A fundamental aim of educational assessment is to maximize validity and reliability in measuring students' abilities (Borsboom et al., 2004). In pursuit of this aim, standards-based educational reform has increased the prevalence of standardized testing and the stakes associated with students' results on these tests (Pellegrino, 2001). In fact, internationally, schools are often funded and publicly ranked based on these results. Yet, the extent to which these tests accurately index students' competencies has been questioned (Pellegrino, 2001; Wiliam, 2003). Specifically, it has been argued that many standardized national curriculum assessments may also assess domain-general (i.e., general purpose, content-free) cognitive capacities in the attempt to assess literacy and numeracy knowledge and skills (Willet and Gardiner, 2009). In support of this suggestion, neuroimaging research suggests that even the simplest literacy and numeracy tasks engage domain-general cognitive networks (Baddeley, 2003; Knudsen, 2007). The domain-general resource most commonly implicated in students' performance on standardized assessments is working memory, whose capacity-limited nature constrains the amount of information that concurrently can be activated, maintained, and manipulated in mind (Engle, 2010). It is therefore unclear whether standardized assessment results reflect students' true literacy and numeracy competencies or whether their scores have been restricted by the limits of their domain-general cognitive resources (e.g., the cognitive demands of the assessment outpacing students' available working memory capacity).

The effects of divergent domain-general cognitive demands are evidenced by research indicating that children's ability to demonstrate their knowledge and skills varies by type of assessment. For instance, in the area of literacy assessment, a recent study found that 75% of students were better able to spell dictated words than correct visually presented misspelled words (the latter based on Australia's National Assessment Program – Literacy and Numeracy, or NAPLAN, method of spelling assessment; Willet and Gardiner, 2009). This finding is consistent with additional studies suggesting that error correction and proofreading tasks typically involve more than just spelling ability (Croft, 1982; Frisbie and Cantor, 1995; although for conflicting results,

*Corresponding author. Tel: +61-2-4221-5165.
E-mail addresses: stevenh@uow.edu.au (S. J. Howard), hana.burianova@cai.uq.edu.au (H. Burianová), john.ehrich@monash.edu.au (J. Ehrich), lkervin@uow.edu.au (L. Kervin), amc998@uowmail.edu.au (A. Calleia), ebarkus@uow.edu.au (E. Barkus), john.carmody@sesiahs.health.nsw.gov.au (J. Carmody), stephen.humphry@uwa.edu.au (S. Humphry).

*Abbreviations:* BOLD, blood oxygenation level-dependent; fMRI, functional magnetic resonance imaging; LV, latent variables; MI, misspelled identified; MNI, Montreal Neurological Institute; MU, misspelled unidentified; NAPLAN, National Assessment Program – Literacy and Numeracy; PLS, Partial Least Squares.

http://dx.doi.org/10.1016/j.neuroscience.2015.03.047
0306-4522/

see Westwood, 1999). This suggests at least some variability in spelling performance may be related to individual differences in domain-general cognitive abilities. Specifically, correcting misspelled words may also require the cognitive flexibility to switch between orthographic representations, thereby placing greater demands on working memory. In fact, working memory has been shown to underlie performance on a broad range of standardized and educational assessments (Gathercole et al., 2003; Strattman and Hodson, 2005; Alloway and Gregory, 2012) and is a particularly powerful predictor of academic achievement (including literacy and numeracy achievement; Blair and Razza, 2007; Best et al., 2011).

Cognitive load researchers have similarly highlighted how the complexity of information and its method of presentation can overwhelm children's limited working memory capacity (van Merrienboer and Sweller, 2005; Kirschner et al., 2011), thus restricting students' ability to acquire and demonstrate their emerging academic competencies. Although fundamentally a theory of learning and instructional design, Cognitive Load Theory principles are similarly applicable to educational assessment in that assessment, like instruction, can impose more or less demand (cognitive load) on test-takers' working memory. Differences in cognitive load across assessments can occur as a function of the inherent complexity of the knowledge and skills being assessed (intrinsic load), immaterial aspects of the assessment relative to the knowledge and skills being assessed (extraneous load), and the mental effort expended on assessment-relevant processes (germane load). For instance, the assessment of whether a student can spell a particular word can be described as being low in element interactivity (successful performance requires minimal reference to, or interaction of, other learned concepts or procedures; Sweller, 1994) compared to correcting a misspelling of that same word. The latter imposes a higher cognitive load, although the specific type of load imposed is less clear. That is, if the assessment aimed to evaluate students' proofreading abilities, the additional load could be characterized as intrinsic load (although this would be an assessment of, at least partly, different knowledge and skills than spelling). However, if the assessment aimed to measure the level of complexity at which students could accurately spell, the additional load could be characterized as extraneous (in that proofreading is a non-essential process for producing the correct spelling of a word). More than just semantics, it is notable that many large-scale, national assessment programs characterize the knowledge and skills they assess using identical terms (e.g., 'spelling'), yet assess these abilities in a highly disparate manner. As a consequence, these assessments may vary in the cognitive demands placed on test-takers' working memory, even when the domain-specific knowledge and skills they assess remain constant. This has important implications for interpretation of assessment results (especially given individual differences in working memory capacity and the resulting differential effect on test performance that may occur) and designing appropriate educational experiences to foster the assessed knowledge and skills.

Although this issue of the domain-general demands of domain-specific assessments is derived from education, it is not easily addressed by traditional educational research methods. For example, neither qualitative nor behavioral studies of spelling assessment are able to conclusively determine the extent to which observed performance differences are spurious (e.g., due to situational or motivational factors), transitory (e.g., due to temporary practice effects), or the product of more fundamental cognitive processes underlying learning and performance (e.g., the varied cognitive load of different modes of assessment). This is an ongoing issue for educational psychologists. Mechanisms of learning and performance are too often defined in operationist terms as psychometric constructs measured exclusively by tests (Michell, 2005; Kelly, 2011), which often are not founded upon substantive theory or an understanding of the function of the brain. The emerging field of educational neuroscience, in contrast, seeks to leverage insights from education, psychology and neuroscience to bridge the gap between the conscious mind and living brain (Szucs and Goswami, 2007). One advantage of applying neuroscientific methods to educational issues is that the contributions of individual neural systems to academic achievement (including domain-general systems) can be identified and quantified (Vander Wyk and Pelphrey, 2011). These neuroanatomical findings can reconcile emerging brain-based insights (such as brain-based evidence of the cognitive load of different forms of assessment) with established educational theory (such as Cognitive Load Theory) to support, refine or advance long-regarded principles of educational best practice (e.g., Whelan, 2007).

The current study sought to combine neuroscientific and behavioral research methods to examine the extent to which domain-general neural correlates contribute to performance on different modes of assessment. Specifically, functional magnetic resonance imaging (fMRI) was used to identify and quantify the domain-general contributions facilitating performance on three different spelling assessments (adapted from Australia's NAPLAN tests, the UK's National Curriculum Tests, and commercial standardized spelling assessments). In addition, out-of-scanner spelling assessments were used to further investigate the relationship between brain (i.e., domain-general neural networks) and behavior (i.e., assessment results). It was expected that triangulation of these results would provide neurological and behavioral evidence that spelling assessments differ in the cognitive load they place on test-takers, as evidenced by: (a) decreased spelling performance on assessments that are higher in cognitive load; and (b) working memory accounting for important variance on assessments that impose greater cognitive load. Specifically, it was expected that error correction methods of spelling assessment (i.e., identify and correct a misspelled word, in line with NAPLAN's method of spelling assessment) would impose greater cognitive load on test takers than dictation forms of assessment (i.e., spell the dictated word, in line with the UK's National Curriculum Tests). As a consequence of