

# UC Santa Barbara

## UC Santa Barbara Previously Published Works

### Title

Validity evidence for a formative writing engagement assessment in elementary grades

### Permalink

<https://escholarship.org/uc/item/7v4719g8>

### Journal

Assessment in Education Principles Policy and Practice, 29(2)

### ISSN

0969-594X

### Authors

Rogers, Paul M  
Marine, Jonathan M  
Ives, Samantha T  
[et al.](#)

### Publication Date

2022-03-04

### DOI

10.1080/0969594x.2022.2054942

Peer reviewed

## Validity evidence for a formative writing engagement assessment in elementary grades

Paul M. Rogers, Jonathan M. Marine, Samantha T. Ives, Seth A. Parsons, Ashlee Horton & Chase Young

To cite this article: Paul M. Rogers, Jonathan M. Marine, Samantha T. Ives, Seth A. Parsons, Ashlee Horton & Chase Young (2022): Validity evidence for a formative writing engagement assessment in elementary grades, *Assessment in Education: Principles, Policy & Practice*, DOI: [10.1080/0969594X.2022.2054942](https://doi.org/10.1080/0969594X.2022.2054942)

To link to this article: <https://doi.org/10.1080/0969594X.2022.2054942>



Published online: 28 Mar 2022.



Submit your article to this journal [↗](#)



View related articles [↗](#)



View Crossmark data [↗](#)



## Validity evidence for a formative writing engagement assessment in elementary grades

Paul M. Rogers<sup>a</sup>, Jonathan M. Marine<sup>b</sup>, Samantha T. Ives<sup>c</sup>, Seth A. Parsons<sup>c</sup>,  
Ashlee Horton<sup>d</sup> and Chase Young<sup>e</sup>

<sup>a</sup>Writing Program, University of California, Santa Barbara, California, USA; <sup>b</sup>English, George Mason University, Fairfax, Virginia, USA; <sup>c</sup>Education, George Mason University, Fairfax, Virginia, USA; <sup>d</sup>Columbia County Schools, Evans, USA; <sup>e</sup>School of Teaching and Learning, Sam Houston State University, Houston, USA

### ABSTRACT

This article reports on the implementation of a formative assessment tool (the Writing Engagement Scale, or WES) in grades 3–5 in schools in the United States. We used confirmatory factor analysis (CFA) to collect validity evidence for the WES for our population. Results demonstrated acceptable validity and reliability. In addition, survey results indicated that teachers perceived the WES to be useful as a formative writing assessment. We make the case that the WES provides an opportunity to inform teachers' practice and help researchers understand the dimensions of students' engagement in writing.

### ARTICLE HISTORY

Received 01 May 2021  
Accepted 03 March 2022

### KEYWORDS

Writing engagement;  
validity; student self-  
assessment; formative  
assessment; elementary  
school

Previous research has shown that writing instruction receives attention in many places around the world – though that attention is context specific, uneven, and in many cases insufficient (Applebee & Langer, 2006; Graham & Rijlaarsdam, 2016). An important element of these instructional contexts are practices associated with writing assessment, which play a determinative role in informing the interactions that occur between teachers, students, and curriculum. As Hillocks (2003) described:

It [assessment] stipulates the kinds of writing that should be taught; it sets the standards for what counts as good writing; and it sets the conditions under which students must demonstrate their proficiency, and, as a result, sets out what students learn. (p. 64)

Many contemporary assessments that attempt to monitor writing proficiency, particularly large-scale standardised writing tests, while having some positive effects, including drawing attention to writing in schools, are somewhat limited in their conceptions of writing and have little to no direct impact on instructional practice or individual student learning. Further, this manner of assessing writing can lead to narrow and restrictive forms of instruction (i.e. teaching to the test), which can potentially cause 'negative impacts on a child's self-efficacy, self-concept' (Behizideh, 2014, p. 125) and motivation (Amrein-Beardsley & Berliner, 2003, p. 32).

**CONTACT** Paul M. Rogers ✉ [paulrogers@writing.ucsb.edu](mailto:paulrogers@writing.ucsb.edu) 📧 Writing Program, University of California Santa Barbara, South Hall, Santa Barbara, CA. 93106, United States

© 2022 Informa UK Limited, trading as Taylor & Francis Group

The history of writing assessment includes many efforts to reframe the concerns of assessment beyond validity and reliability (Yancey, 1999) towards assessments with practical instructional value. Previous research has indicated how much of the early work on assessment focused primarily on reliability (Huot, 1990), because reliable tests were easier and less costly to administer (Yancey, 1999). While reliability remains a pre-eminent concern in assessment research to this day, equally important to an assessment's validity and reliability are teachers' confidence that it generates data, information, or scores 'upon which instructional decisions can be based' (G. Brown & Harris, 2014).

Accordingly, the need to develop productive, positive, and scalable writing assessments presents an important opportunity to align what we can accurately measure with what teachers can actionably enact in the classroom with what we have learned through research about learning to write and writing instruction. Instructionally validated and psychometrically tested assessments are particularly important in the elementary grades to guide daily instruction and in doing so contribute to the growth of knowledge concerning writing assessment for young children while supporting the development of more playful, effective, and engaged writers.

Given the potential positive impact of assessment on learning and instruction, we set out to design a formative self-assessment tool drawing on Moffett's (1992) fundamental premise that assessment must take place at the site of in situ learning already occurring in the classroom: 'the most efficient education would never require of learners that they would do anything especially for evaluations that they would not be doing anyway in order to learn' (p. 2). Our goal was to design a student self-assessment that could function as a natural part of a teacher's lesson design and assignment sequencing and that would provide enough flexibility to work across genres in elementary grade levels (2–5). From the beginning, we set out to design a self-assessment tool with age-appropriate instructions (Panadero et al., 2016; G. T. L. Brown & Harris, 2013) that students could learn to use quickly and easily, and that would provide teachers with actionable information that they could use to inform their instruction and that would provide insight into multiple dimensions of their students' writing processes and their engagement with writing.

The current paper presents our work in exploring validity properties of this tool and describes how its results have the potential to support teachers and serve students by changing writing instruction at the individual and whole class level. The tool is a student self-assessment – the Writing Engagement Scale (WES) – a sixteen-item scale (see Appendix for complete scale) that measures elementary students' perceptions of their text production and participation in writing tasks. The primary purpose of the tool is to provide teachers with valuable information for formative assessment at the level of the whole class and for individual student instruction. Thus, the WES includes items related to the affective, behavioural, cognitive, and social dimensions of writing, which comprise our nascent model of writing engagement.

## Related literature

### *Self-assessment and formative assessment*

While there are ongoing debates regarding the relationship of assessment to learning (Baird et al., 2017; Black & William, 2018; William, 2017), arguably, a great deal of progress has been made in pushing forward work associated with formative assessment; that is, assessments that are used to provide information that teachers can use to adapt instruction and better support student learning. ‘Whether it is a curriculum or student achievement that is being evaluated, the evaluation is formative if the information generated is used to make changes to what would have happened in the absence of such information’ (William, 2006, p. 264). In designing formative assessments, those responsible for instruction can draw on assessment data from a wide variety of sources including self-assessment data (Calfee et al., 2014). Self-assessment is also a vital part of self-regulation, as it allows the student to draw on important metacognitive competencies (e.g. self-observation, self-judgement, self-reaction, task analysis, self-motivation, and self-control; Kitsantas & Zimmerman, 2006). Thus, as Brookhart (2016) explains, one of the main functions of self-assessment is to engage students at a deep level with their learning goals and the criteria for success, a foundational principle in formative assessment. Most importantly, what categorises an assessment as formative is its direct impact on learning (Swaffield, 2011; Weeden et al., 2002) and/or instruction (William, 2006).

An assessment of a curriculum is formative if it shapes the development of that curriculum. An assessment of a student is formative if it shapes that student’s learning. Assessments are formative, therefore, if and only if something is contingent on their outcome, and the information is actually used to alter what would have happened in the absence of the information. (William, 2006, p. 264)

Among the common procedures used to inform formative assessments are student self-assessments, which have been clearly linked to formative assessment (Andrade et al., 2010; Berry, 2011; Black & William, 1998; Leahy & William, 2012), and can involve both description (i.e. these are the characteristics of my work) and evaluation (i.e. this is how good my work is and what it is worth). A number of studies have demonstrated a positive association between self-assessment, learning, and achievement (see, G. Brown & Harris, 2014, for a review). Self-assessment encourages ‘ownership of learning and the use of self-regulatory strategies’ (Panadero et al., p. 804) and is also positively associated with student empowerment in the assessment process (Black & William, 1998; Nicol & Macfarlane-Dick, 2006; Tan, 2012; Taras, 2010).

While there is little scholarly consensus on the typologies of student self-assessment, Panadero et al. (2016) taxonomized self-assessment using five criteria (knowledge interest, involvement, power and transparency, presence, and format) to generate 21 categories. Citing Tan (2001), one important category of self-assessment was self-awareness, where ‘students are aware of their thinking processes and assess them, but without a formal comparison (i.e. no external criteria or teachers)’ (Panadero et al., 2016, p. 806). Another equally important category of self-assessment was self-rating, where a rating system (e.g. smiley face ordinal rankings) is used by students ‘to judge the quality or quantity aspects of their work’ (Panadero et al., 2016, p. 809). Both these categories are instructive in conceptualising the design and implementation of self-assessment tools,

which Panadero et al. (2016) contend can increase self-regulated learning strategies (Kostons et al., 2012), enhance students' self-efficacy (Olina & Sullivan, 2004; Ramdass & Zimmerman, 2008), increase learning and academic performance (G. Brown & Harris, 2014), and empower students in the assessment process (Black & Wiliam, 1998; Nicol & Macfarlane-Dick, 2006; Tan, 2012; Taras, 2010). What all of this research indicates is the potential benefit of self-assessment for the purposes of formative assessment. We designed the WES to help teachers more dependably interpret student writing performances by providing actionable information to them with the goal of indexing the behaviours associated with supporting more engaged writers.

## **Engagement**

There is a growing international research interest in student engagement and its relationship to learning (Fredricks et al., 2004; Harris, 2011; Lee et al., 2021; Pino-James & Nicolás Pino-James, 2018; Parsons et al. 2018). Research suggests that engagement is a reliable predictor of future academic performance and is associated with long-term positive outcomes (Fredricks et al., 2004; Shernoff, 2013; Upadhyaya & Salmela-Aro, 2013). As Christenson et al. (2012) succinctly stated, 'Student engagement drives learning' (p. 817).

Research on literacy engagement, and particularly in reading engagement, has similarly shown a positive impact on student performance (Alexander, 2018; Guthrie et al., 2012; Lee et al., 2021; Authors et al., 2018). While research on educational and literacy engagement are both increasing, research on writing engagement is much less developed, and is also frequently associated (or conflated) with writing motivation (see, Camacho et al., 2021 for a review). Nonetheless, writing researchers have sought to investigate the relationship between writing engagement and feedback (Schildkamp et al., 2020; Wolsey & Grisham, 2007), social learning strategies (Farizka & Cahyono, 2021; Liu et al., 2018), task relevance (Hawthorne, 2008), writing prompts (Cho, 2019), and perceptions of self-efficacy (Kitsantas & Zimmerman, 2006; Tsao, 2021). These studies all conceive of writing engagement as a multidimensional construct, typically composed of affective (Boscolo & Hidi, 2007), behavioural (Cho, 2019; Liu et al., 2018), cognitive (Kitsantas & Zimmerman, 2006), and social components (Farizka & Cahyono, 2021) in order to 'draw attention to the importance of the individual's conscious, motivated, and aware engagement in learning strategies' (Brown, 2015).

Similarly, we present writing engagement as a multidimensional construct including affective, behavioural, cognitive, and social components that influences one's intentional and thoughtful involvement in writing-related activity (Finn & Zimmer, 2012; Fredricks et al., 2004). Behavioural engagement is participation, effort, and persistence; affective engagement is related to interest and enjoyment; and cognitive engagement includes strategy use and self-regulation (Fredricks & McColskey, 2012; Reschly & Christenson, 2012). More recently, scholars have included a fourth dimension, social engagement, presented as the degree to which students collaborate and interact with others (Alexander, 2018; Ivey & Johnston, 2013, 2015).

Our construct of writing engagement is devised by subject-matter experts in writing and literacy (Baird et al., 2017) and supported by the dominant views of learning and writing as social processes (Graham & Harris, 2013; National Academies of Sciences,

Engineering and Medicine, 2018; Vygotsky, 1978). Student engagement is associated with enhanced academic performance (Ivey & Johnston, 2013), and more time spent writing enhances students' writing ability (Authors, 2008; Graham, 2019). 'Additionally, when students are engaged they are more reflective and seek out more support for their writing, which increases their opportunities to learn' (Rogers, 2008, p. 181). Further, we support the view that engagement enhances learning and can also be *the outcome of success in learning* and not merely a precursor that causes learning. As Berninger and Hidi (2007) suggest, 'the more successful a student is in the learning process, the more engaged and the less avoidant the student becomes in future learning activities, thus increasing the likelihood that the success will persist' (p. 167). Unfortunately, researchers have typically found that students' engagement tends to decline as they advance through the grade levels (Skinner & Pitzer, 2012).

We created the WES to assess multiple dimensions of writing engagement: affective, behavioural, cognitive, and social writing engagement dimensions (Alexander, 2018; Fredricks et al., 2004). The multidimensionality of this tool provides a robust snapshot of student writing engagement that is not afforded by existing measures. This tool was designed as a formative classroom assessment in which students self-assess their experience with a particular writing assignment. Throughout the scale design process, we used student and teacher feedback to modify the scale to ensure it was classroom-friendly and useful for instruction. The first step in determining the WES's usefulness is to evaluate its validity and reliability, which we do in the current study guided by the following research questions:

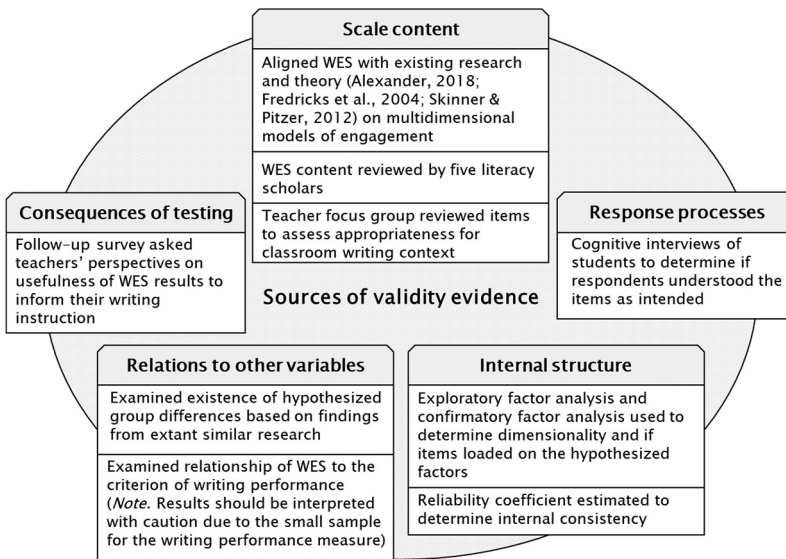
1. Does the WES provide valid and reliable scores of upper elementary student writing engagement?
2. To what extent do teachers' perceptions of the usefulness of the WES support its validity?

## Method

The current study addressed the research questions by examining the validity evidence of the WES, a questionnaire developed for this study, in multiple classrooms (grades 3–5) in the southern United States. The study was approved by George Mason University's Office of Research Integrity and Assurance IRBNet number: 1,698,770–1. Our understanding of validity is aligned with *The Standards for Educational and Psychological Testing* (AERA, NCME, APA, 2014). In [Figure 1](#) we present the validity evidence we collected and its alignment with the sources of validity evidence outlined by the standards.

### *Scale development and pilot study*

We created the WES to measure students' writing engagement in the situated context of a classroom-based assignment. This scale was created as a formative assessment of student writing, and unlike typical summative writing assessments, the WES is not exclusively focused on performance. Rather, it is a self-assessment of students' engagement in a writing task. We conceptualised this scale based on previous models of literacy



**Figure 1.** Types of validity evidence collected in standards framework. Note. *Standards* refers to the *Standards for Educational and Psychological Testing* (AERA, APA, and MCME, 2014).

engagement, which include affective, behavioural, cognitive, and social dimensions (Alexander, 2018; Parsons et al., 2015). As such, we devised items to reflect each of the four proposed dimensions.

To acquire evidence of validity, we asked five literacy scholars to review the scale to evaluate its alignment with our purposes. As one of our goals for the WES was to ensure ease of administration and helpful formative information for teachers, we used a focus group format to share the scale with two elementary teachers and asked for their feedback about its potential alignment with their teaching practices and students; additionally, we conducted cognitive interviews with school-age children. The WES, then, includes items based on scholar, teacher, and student input.

Before conducting our study, we carried out a pilot study with 179 students in grades 2–5 and seven teachers. We used exploratory factor analysis to determine the number of factors and evaluate the internal structure of the WES. We refined the WES based on these results.

## Current study

### Sample

We received 330 responses to the WES (see, Table 1 for item text and additional item information). Prior to our quantitative analysis, we calculated Mahalanobis distance to identify multivariate outliers that would skew the results. From this analysis, we eliminated three outliers. Our final sample of 327 included 156 male students, 163 female students, and 8 students who did not identify their gender. Regarding grade level, the sample consisted of 84 students in third grade (25.69%), 136 students in fourth grade (41.59%), and 107 students in fifth grade (32.72%). Concerning school demographics,



**Table 1.** Revised WES item factor, item order, and text.

Factor/ order	Item text
AE1 <sup>a</sup>	When working on this writing assignment, I was interested in what I was writing.
AE2 <sup>b</sup>	Working on this writing assignment was boring.
AE3 <sup>c</sup>	When working on this writing assignment, I felt good.
AE4 <sup>b</sup>	I would like to complete a writing assignment like this again.
BE1 <sup>b</sup>	I got distracted when I was working on this writing assignment.
BE2 <sup>b</sup>	When working on this writing assignment, I just wanted to get it finished.
BE3 <sup>c</sup>	I tried hard to do well on this writing assignment.
BE4 <sup>c</sup>	I worked as hard as I could on this writing assignment.
CE1 <sup>a</sup>	When working on this writing assignment, I created a plan before I started writing.
CE2 <sup>a</sup>	When working on this writing assignment, I reread to see if I could make it better.
CE3 <sup>b</sup>	When working on this writing assignment, I thought carefully about the words I used.
CE4 <sup>d</sup>	I asked myself questions as I was writing to make sure my writing made sense.
SE1 <sup>a</sup>	When working on this writing assignment, I talked with other students about my writing.
SE2 <sup>b</sup>	When I was finished with this writing assignment, I wanted to share it with others.
SE3 <sup>b</sup>	I don't like it when classmates or teachers try to improve my writing.
SE4 <sup>b</sup>	I can think of at least one person who would want to read this writing.

Items AE2, BE1, BE2, and SE3 were reverse coded. AE = Affective Engagement; BE = Behavioural Engagement; CE = Cognitive Engagement; SE = Social Engagement.

<sup>a</sup>Adapted from original WES.

<sup>b</sup>New item created by authors.

<sup>c</sup>Item adapted from Skinner et al. (2009).

<sup>d</sup>Item adapted from Wang et al. (2014).

301 students were from a school with the following demographics: 76% White, 10% Black, 6% Hispanic, 4% Asian, and 4% multiracial. Approximately 8% of students were economically disadvantaged. Twenty-six students were from a school without available demographic information.

### Data analysis

We used SPSS Version 28 to analyse descriptive statistics, examine the linearity and normality of our data, create the averaged composites of variables after the confirmatory factor analysis (CFA), determine group differences, and obtain reliability information in the form of Cronbach's alpha. See, Table 2 for item correlations and descriptives. As part of the implementation of the WES, we were able to collect some student writing samples. Unfortunately, we only received writing samples from 15 students, seven girls and eight boys (third grade  $n = 2$ , fourth grade  $n = 4$ , fifth grade  $n = 8$ ). Although we caution placing too much emphasis on the results, we carried out our plan to analyse the correlations between students' writing engagement and their writing quality. Writing quality was measured using the Analytic Writing Continuum (AWC; National Writing Project, 2010); two authors scored student essays on the AWC's writing performance attributes of content, conventions, diction, sentence fluency, stance, and structure. Additionally, we recorded the number of sentences in each writing sample. The authors came together to discuss any confusions or disagreements and came to an agreement about each essay's attribute scores.

**Table 2.** Writing engagement scale item correlations and descriptives.

Item	AE1	AE2	AE3	AE4	BE1	BE2	BE3	BE4	CE1	CE2	CE3	CE4	SE1	SE2	SE3	SE4
AE1	4.05 (0.95)															
AE2	0.40	3.88 (1.11)														
AE3	0.41	0.50	4.00 (0.90)													
AE4	0.34	0.46	0.45	3.74 (1.14)												
BE1	0.27	0.19	0.21	0.17	3.46 (1.16)											
BE2	0.26	0.27	0.28	0.41	0.16	3.24 (1.28)										
BE3	0.35	0.26	0.37	0.33	0.19	0.22	4.39 (0.75)									
BE4	0.34	0.33	0.33	0.29	0.19	0.29	0.55	4.23 (0.81)								
CE1	0.32	0.30	0.30	0.31	0.23	0.16	0.39	0.22	3.70 (1.10)							
CE2	0.28	0.28	0.31	0.30	0.20	0.20	0.42	0.38	0.38	4.05 (0.91)						
CE3	0.34	0.35	0.34	0.27	0.15	0.17	0.40	0.27	0.32	0.36	4.12 (0.78)					
CE4	0.35	0.28	0.31	0.29	0.12	0.17	0.32	0.30	0.20	0.45	0.36	3.82 (0.94)				
SE1	0.25	0.20	0.16	0.22	0.07	0.16	0.23	0.18	0.23	0.22	0.26	0.15	3.14 (1.22)			
SE2	0.30	0.41	0.31	0.40	0.25	0.31	0.39	0.34	0.37	0.39	0.34	0.33	0.33	3.49 (1.24)		
SE3	0.15	0.24	0.16	0.19	0.18	0.19	0.17	0.07	0.18	0.16	0.21	0.20	0.15	0.19	3.53 (1.19)	
SE4	0.28	0.19	0.18	0.21	0.20	0.23	0.21	0.25	0.16	0.19	0.12	0.12	0.26	0.32	0.09	3.85 (1.09)

Item mean and parenthetical standard deviations are presented along the diagonal. *r* coefficients > .10 are significant at a *p* < .05 level.

Based on reviewer feedback, we used CFA to determine the structural validity of the WES. The CFA was conducted with the lavaan package (Rosseel, 2012) and estimation of an omega coefficient as an additional method of assessing reliability was completed using the semTools package (Jorgensen et al., 2021) in R version 4.0.3. Examination of histograms revealed the WES data were non-normal; consequently, we used one of lavaan's robust maximum likelihood estimators, MLM, which utilises robust standard errors and a scaled chi-square statistic (Satorra & Bentler, 1994) to provide estimates that are robust to non-normal data.

To evaluate CFA model fit, we examined the Satorra-Bentler scaled chi-square ( $\chi^2_{SB}$ ), the robust Comparative Fit Index ( $CFI_{SB}$ ), the robust Root Mean Square Error of Approximation ( $RMSEA_{SB}$ ), and the Standardised Root Mean Square Residual (SRMR). We used Hu and Bentler's (1999) recommendations that global fit cut-off values close to  $CFI > .95$ ,  $RMSEA < .06$ , and  $SRMR < .08$  indicate model fit. To compare non-nested models, which are models with different observed variables, we looked at Akaike information criterion (AIC) and Bayesian information criterion (BIC) values. Lower AIC and BIC values indicate better fit (Brown, 2015); based on modifications made to the models, we expected AIC and BIC values to decrease. We examined standardised factor loadings, standardised residuals, and modification indices, in addition to our knowledge of writing and engagement, to make model modifications. Factor loadings, standardised residuals, and modification indices are measures of local fit, as compared with the global fit indices (i.e. CFI, RMSEA, SRMR, AIC, and BIC) explained above. Brown (2015) noted for applied research, standardised factor loadings greater than .30 to .40 are considered meaningful, although Kline (2016) posits standardised factor loadings should be greater than .70 because this indicates about 50% of the variance in the item is explained by the factor. Large standardised item residuals (i.e. greater than  $|2.58|$ ) and weak standardised factor loadings were considered indicators of local model misfit (Brown, 2015). These were used, in addition to consideration of modification indices and theoretical/conceptual plausibility, to make decisions about model modifications.

Our hypothesised model included the four factors of affective engagement, behavioural engagement, cognitive engagement, and social engagement with four indicators on each factor. We interpreted evidence of convergent validity through the strength and significance of factor loadings. Strong factor loadings indicate variance in the items is explained by the factor. For example, for an item with a standardised loading of .75 on the affective engagement factor, about 56% of the item variance is explained by the latent affective engagement factor and 44% of the item variance is attributed to error. We examined the factor correlations to gather discriminant validity evidence suggesting the four hypothesised types of writing engagement are indeed separate. Factor correlations equal to or greater than .85 were considered indicative of poor discriminant validity (Tabachnick et al., 2019). If factor correlations are that high, researchers should consider collapsing factors due to the lack of evidence that the specified factors measure different latent constructs (Brown, 2015).

## Results

### Research question 1

In our CFA model, we specified four indicators on each of the four hypothesised engagement factors of affective, behavioural, cognitive, and social writing engagement. According to the selected global fit indices, this model fit the data (see, Table 3 for the global model fit indices for each four-factor model tested). Because our model fit the data globally, we moved on to identifying local areas of misfit (see, Table 4 for local model fit information and respecification decisions). Residuals and low standardised factor loadings suggested problems with the reverse-worded items on the WES. In accordance with recommendations (Brown, 2015), we specified a model with a negative method effect factor connected to the negatively worded indicators. None of the indicators had significant loadings on the negative method effect factor, so we did not pursue this model further. Instead, we removed an item before examining the next model. We continued to run each model, consider global fit, examine local fit, and make model specifications until we were confident in the integrity of the model. See, Figure 2 for standardised factor loadings and factor correlations of our final model.

In terms of convergent validity, conceptualised through a CFA framework as the strength of the relationship between the factors and the items, our findings are mixed. All of the items loaded significantly on their hypothesised factor. However, the ideal magnitude of factor loadings is greater than .70. Three of our items met this criteria, but nine items did not. The items in the behavioural engagement factor, which is our strongest performing factor in terms of factor loadings, were both greater than .70, which means the majority of the variance in items is explained by the behavioural engagement factor. The items specified to the affective engagement and cognitive engagement came close to meeting the greater than .70 threshold, but more variance in the items is attributed to error than to the factors. Our social engagement factor fared the worst in terms of factor loadings, although a strength is that about 58% of the variance of one item, ‘When I was finished with this writing assignment, I wanted to share it with others’, was explained by the social engagement factor.

**Table 3.** Global goodness-of-fit indices for four-factor writing engagement models.

Model	$\chi^2_{SB}$		CFI <sub>SB</sub>	RMSEA <sub>SB</sub>		SRMR	AIC	BIC	$\Delta$ AIC	$\Delta$ BIC
	Value	df		Value	90% CI					
Model 1 <sup>a</sup>	146.29*	98	.95	.04	[.03, .06]	.05	13840.30	13984.31		
Model 2 <sup>b</sup>	126.11*	84	.96	.04	[.03, .06]	.05	12820.77	12957.21	-1019.53	-1027.10
Model 3 <sup>c</sup>	112.06*	71	.96	.04	[.03, .06]	.05	11822.90	11951.76	-997.87	-1005.45
Model 4 <sup>d</sup>	76.92	59	.98	.03	[.00, .05]	.04	10775.94	10897.22	-1046.96	-1054.54
Model 5 <sup>e</sup>	56.76	48	.99	.03	[.00, .05]	.04	9869.60	9983.30	-906.34	-913.92

$\chi^2_{SB}$  = Satorra-Bentler (SB) adjusted chi-square; CFI<sub>SB</sub> = the SB scaled comparative fit index; RMSEA<sub>SB</sub> = the SB scaled root mean square error of approximation; SRMR = standardised root mean square residual; AIC = Akaike information criterion; BIC = Bayesian information criterion.

<sup>a</sup>In Model 1, 16 items of the revised WES loaded onto the four factors of affective, behavioural, cognitive, and social writing engagement (four items loaded on each factor).

<sup>b</sup>In Model 2, one item (SE3) was removed from the analysis.

<sup>c</sup>In Model 3, a second item (BE1) was removed from the analysis.

<sup>d</sup>In Model 4, a third item (BE2) was removed from the analysis.

<sup>e</sup>In Model 5, a fourth item (CE1) was removed from the analysis.

\*  $p < .05$

**Table 4.** Confirmatory factor analysis model local areas of fit and respecification decisions.

Model	Number of indicators on each factor				Standardised factor loadings	High residuals <sup>a</sup>	Item removed	Reason for removal
	AE	BE	CE	SE				
Model 1	4	4	4	4	.71 – .31	5, (–3.43–3.99)	–	–
Model 2	4	4	4	3	.73 – .33	5, (–3.38–4.00)	SE3	Low loading (.31) Conceptually inconsistent <sup>b</sup>
Model 3	4	3	4	3	.74 – .41	5, (–3.65–4.20)	BE1	Low loading (.33) Modification indices <sup>c</sup>
Model 4	4	2	4	3	.80 – .43	2, (–3.00 – –2.69)	BE2	Low loading (.41) Modification indices <sup>c</sup>
Model 5	4	2	3	3	.78 – .43	1, (–2.62 <sup>d</sup> )	CE1	High residuals <sup>a</sup> Conceptually inconsistent <sup>e</sup>

The item removed column identifies the item that was removed before the model identified in the stub column was run.

AE = Affective Engagement; BE = Behavioural Engagement; CE = Cognitive Engagement; SE = Social Engagement.

<sup>a</sup>Standardised residual values > |2.58|.

<sup>b</sup>This item is more general than the other items on the WES; it is the only item that is not about the specific writing assignment.

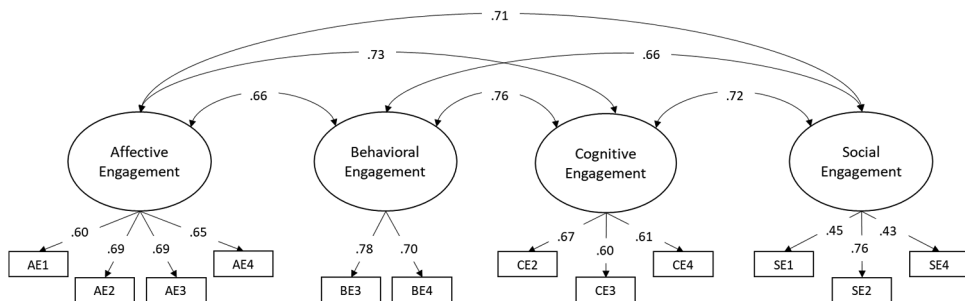
<sup>c</sup>Lack of theoretically-plausible alternatives based on modification indices.

<sup>d</sup>This is residual between the affective engagement item, 'Working on this writing assignment was boring', and the behavioural engagement item, 'I tried hard to do well on this writing assignment'. The negative standardised residual suggests the relationship between these two items was overestimated. Although the standardised residual was over our threshold, we declined to make further changes to the model because despite the high residual, both of these indicators had high loadings on their respective factors and we could not conceptually justify the removal of either item.

<sup>e</sup>This was the only item to measure cognitive engagement in the pre-writing process rather than the more metacognitive in-the-moment nature of the remaining three cognitive engagement items.

The model demonstrated discriminant validity, which in the CFA framework means the factor correlations indicate there are indeed four separate factors. Our factor correlations ranged from  $r = .66$  to  $r = .76$ , which is below the suggested discriminant validity threshold of  $r = .85$ . To further test our four-factor model, we examined a unidimensional model in which one factor, writing engagement, informed our 12 remaining indicators.

*Final Four-Factor Writing Engagement Model*



**Figure 2.** Final four-factor writing engagement model. Note. All coefficients are significant ( $p < .05$ ) standardised values.

This model did not fit the data well,  $\chi^2_{SB}(54) = 127.61, p = .00$ ; CFI<sub>SB</sub> = .90; RMSEA<sub>SB</sub> = .07 [.06, .09]; SRMR = .05; AIC = 9947.32; BIC = 10,038.28. We found further evidence of model misfit when examining the standardised residuals, seven of which were greater than 2.58, and ranged from -3.78 to 4.08. Standardised factor loadings of the unidimensional model ranged from .36 to .63. Because the unidimensional model is nested in the final four-factor model (Model 5), which means the same observed variables are used in both models, we can statistically compare the unidimensional model to the final four-factor model using a chi-square difference test. The results of the scaled chi-square difference test (Satorra & Bentler, 2001) revealed Model 5 is significantly better than the unidimensional model,  $\Delta \chi^2_{2SB}(6) = 67.15, p = .00$ .

Based on the results of our CFA, we created composite variables of the four types of writing engagement by calculating the average of each factor's items for each student. Student descriptives for each subscale are as follows: affective engagement ( $M = 3.92, SD = 0.77$ ), behavioural engagement ( $M = 4.33, SD = 0.65$ ), cognitive engagement ( $M = 4.01, SD = 0.67$ ), and social engagement ( $M = 3.49, SD = 0.86$ ). We calculated the reliability of each subscale using Cronbach's alpha. According to Sattler and Hoge's (2006) guidelines, our social engagement subscale ( $\alpha = .57$ ) was unreliable, our cognitive engagement subscale ( $\alpha = .66$ ) was marginally reliable, and our affective engagement ( $\alpha = .75$ ) and behavioural engagement ( $\alpha = .70$ ) subscales were relatively reliable. Based on reviewer feedback and current debate in the field criticising the use of Cronbach's alpha (Dunn et al., 2014; McNeish, 2018), as well as researchers' advocacy for the use of coefficient omega in place of Cronbach's alpha (Hayes & Coutts, 2020; McNeish, 2018) we also estimated coefficient omega (Raykov, 2001). Omega values were equal to alpha values for all subscales except for the social engagement subscale ( $\alpha = .57, \omega = .58$ ). On the alternative side of the debate about the utility of Cronbach's alpha, authors explained the anticipated similarity in results between the two estimates (Savalei & Reise, 2019) and are less inclined to disregard alpha altogether (Raykov & Marcoulides, 2019).

To assess evidence for validity based on relations to other variables, we examined correlations between writing engagement subscale scores, six writing performance attributes, and the number of sentences written. Although examining relations between the focal variables (i.e. writing engagement subscales) and related criterion variables is a legitimate way to gather evidence for validity, our sample size ( $n = 15$ ) for this analysis was limited by external circumstances and as such, we do not believe these results are generalisable nor do they have much statistical power. Despite this, we briefly present our findings from the correlational analysis because we see value in exploring the evidence we have. The cognitive engagement subscale significantly and positively correlated with the writing performance attributes of content ( $r = .58$ ), conventions ( $r = .54$ ), diction ( $r = .62$ ), stance ( $r = .55$ ), and structure ( $r = .61$ ) as well as with the number of sentences written ( $r = .63$ ). The only measure of writing performance it was not significantly associated with was sentence fluency. No other subscales had any significant relationships with the attributes of writing performance nor with number of sentences. We will not spend too much time on this finding, but we believe this is a conceptually reasonable finding. Cognitive engagement in writing is when students think and act strategically

during a writing task. Constructs such as strategic writing behaviours and self-regulated learning strategies are positively related to writing quality (Graham et al., 2017; Regan et al., 2017), and our findings parallel those in extant research.

Our final examination of how the WES functions as a measure of writing engagement in a sample of upper elementary students was to look at group differences in WES scores. We first compared students' writing engagement based on sex using independent t tests. The results of the t tests revealed girls had higher subscale scores for all types of writing engagement. Specifically, girls had higher affective engagement ( $M = 4.07$ ,  $SD = 0.67$ ) than boys ( $M = 3.79$ ,  $SD = 0.82$ ),  $t(299.52) = 3.37$ ,  $p = .00$ , Cohen's  $d = 0.40$ ; higher behavioural engagement ( $M = 4.45$ ,  $SD = 0.56$ ) than boys ( $M = 4.21$ ,  $SD = 0.71$ ),  $t(293.26) = 3.34$ ,  $p = .00$ , Cohen's  $d = 0.38$ ; higher cognitive engagement ( $M = 4.16$ ,  $SD = 0.60$ ) than boys ( $M = 3.84$ ,  $SD = 0.69$ ),  $t(317) = 4.44$ ,  $p = .00$ , Cohen's  $d = 0.50$ ; and finally, girls had higher social engagement ( $M = 3.68$ ,  $SD = 0.78$ ) than boys ( $M = 3.29$ ,  $SD = 0.90$ ),  $t(306.15) = 4.10$ ,  $p = .00$ , Cohen's  $d = 0.46$ . In terms of practical significance, according to Cohen's (1988) guidelines, the differences between boys' and girls' affective and behavioural writing engagement are small to medium and the differences between their cognitive and social writing engagement were medium. Next, we ran a multivariate analysis of variance (MANOVA) to examine differences in the four types of student writing engagement based on grade level. We chose to use a MANOVA because we know the dependent variables (affective, behavioural, cognitive, and social writing engagement) are related, but we also know multicollinearity will not be an issue for this analysis. There was no significant difference in student writing engagement based on grade level,  $\lambda = .97$ ,  $F(8, 642) = 1.33$ ,  $p = .23$ .

In all, we found mixed quantitative evidence in support of the WES's validity as a measure of elementary school student writing engagement. All items, even the ones removed from the analysis, had significant loadings on their hypothesised factors. However, we saw a range of standardised item loadings from .43 to .78. While all of the items had standardised factor loadings above .40, only three items had standardised factor loadings equal to or greater than .70. We have some evidence of convergent validity, but overall the magnitude of factor loadings is a limitation of our study. The social engagement subscale is the weakest part of the WES. Two of its three factor loadings are below .50, indicating a large amount of error. Additionally, this subscale is unreliable according to Sattler and Hoge's (2006) guidelines for Cronbach's alpha. Alternatively, the behavioural engagement subscale, even though it only has two items, displays evidence of convergent validity. Its items have standardised factor loadings greater than or equal to .70, which indicates the majority of the variance in items is explained by the behavioural engagement factor instead of error. Additionally, the behavioural engagement subscale had a relatively reliable Cronbach's alpha. However, the behavioural engagement subscale was reduced to two items, which may not fully capture the breadth of behavioural engagement.

We found strong evidence of discriminant validity in the CFA framework; factor correlations were all smaller than  $r = .85$  (Tabachnick et al., 2019). This means the four factors are divergent enough that they should not be collapsed into one factor or scale. This was an important aspect of validity for the WES. Additionally, our final four-factor model fit the data statistically better than a unidimensional model. Finally, group differences provided some additional evidence for validity. In line with previous research,

we found differences in writing based on sex (Graham et al., 2017), but in contrast with previous research, we did not see significant differences in student engagement based on grade level (Skinner & Pitzer, 2012).

### Research question 2

In an effort to further validate the WES (Walpole et al., 2021), after the administration of the WES we immediately sent all participating teachers the data generated from the students in their classroom and invited them to take a six-question online survey (see, Table 5). Each teacher received an email with the complete results at the whole class and individual student levels. The survey questions asked the teachers to reflect on the usefulness of the WES's results for their instruction, and which item(s) on the scorecard were most useful.

As part of the survey, we asked each instructor to select one student's scorecard results and to look at the student's text submission and to reflect on any insights into the student's writing activity; additionally, we asked teachers to consider the ways in which the results might inform future whole class instruction. Finally, we asked for the teachers' input on the WES itself as a tool and for their suggestions in making it more useful for the purposes of providing formative feedback. Teachers from 18 individual classrooms responded to the survey, of which 16 stated that the WES was useful for improving classroom instruction across a number of different areas, especially for informing instructional planning. The items on the WES which teachers deemed most valuable related to the cognitive dimensions of writing engagement (e.g. 'When writing this piece, I created a plan before I started writing' and 'When writing this piece, I reread to see if I could make it better').

### Discussion

In this project, we attempted to design an assessment tool, the WES, for researchers and teachers to use to assess students' writing engagement. Further, this measure of students' engagement was also designed to be useful for formative assessment purposes and to help researchers and teachers learn more about and better support students' writing development. In our study we gathered validity and reliability information based on a sample of 327 students in grades 3–5. We also gathered teachers' reflections on their students' WES results to ascertain teachers' perceptions of its usefulness as a formative assessment tool. Initial results suggested that the WES demonstrated acceptable psychometric properties. We discuss this result more below. We also discovered that teachers found the WES to be a useful and insightful tool for formative classroom writing assessment.

**Table 5.** Teacher survey items.

Survey item text
In what ways are the overall, class-wide scorecard results useful for you in providing feedback to your students?
Which items in the scorecard offer the most useful information in terms of providing feedback? Pick two.
After selecting one student's scorecard results and looking at their accompanying text, in what ways does the scorecard provide meaningful insight into the student's writing activity?
How do the overall class scores on this assignment help you to think about how you might adjust or adapt your instruction for your next writing assignment?
Please elaborate on any other ways in which the scorecard results provide useful information for you as a teacher.
Is there anything you would change or add to the scorecard to make it more helpful?



## **Validity**

Our CFA findings provided us with information about the validity of the WES as a measure of students' writing engagement. With our CFA we were able to model a social writing engagement factor and establish a multidimensional factor structure, which is aligned with our conceptual framework of engagement having affective, behavioural, cognitive, and social components (Alexander, 2018; Ivey & Johnston, 2013, 2015) and as such provides supportive validity evidence based on internal structure. Our CFA provided a theoretically-justified model of writing engagement. Reliability of the WES was acceptable (except for the social engagement dimension), which also provides positive evidence for validity based on internal structure. Finally, group differences based in gender were aligned with previous research (Graham et al., 2017), supporting our validity evidence based on relations to other variables, although we did not find the expected differences in grade level (Skinner & Pitzer, 2012).

The validity of the WES was also supported by teacher survey feedback. The majority of teacher comments indicate that reviewing the WES results helped them to better understand students' perceptions of their own writing, as well as the behaviours undergirding the production of student writing, i.e. the degree to which they are engaged cognitively, behaviourally, emotionally, and socially with writing – and to more carefully orchestrate their instructional activities to match them within student's zones of proximal development (Vygotsky, 1978). These comments included both validation of effective instructional practices that students took up and identified, as well as areas for future changes in instruction, for example, 'It [helps me to know the areas they may need more writing conferences in [and] helped me to see the areas they struggle most and come up with some mini lesson teaching points'. In our view, the teachers' positive perceptions of the usefulness of the WES for making better informed instructional choices provide another layer of support to psychometric conceptions of the tool's validity.

## **Self-assessment**

Students who used the WES took 'ownership of learning and the use of self-regulatory strategies' in keeping with previous research on the positive associations between student self-assessment and student empowerment (Black & Wiliam, 1998; Nicol & Macfarlane-Dick, 2006; Tan, 2012; Taras, 2010). While a positive association between self-assessment and learning and achievement, as previous research has shown, has yet to be established with this tool, the WES is nonetheless encouraging self-awareness, defined by Tan (2012) as when 'students are aware of their thinking processes and assess them, but without a formal comparison (i.e. no external criteria or teachers)' (p. 806). The WES is also a self-rating self-assessment, because a rating system (e.g. ordinal ranks) was used by students 'to judge the quality or quantity aspects of their work' (p. 809). The WES comports to Tan's (2012) taxonomy of self-assessment types and lends further credence to the link between self and formative assessment, as our results clearly indicate the formative potential of students' self-assessment results.

## ***Writing engagement***

We found support for the WES's alignment with conceptualisations of engagement as a construct that includes affective, behavioural, and cognitive components (Fredricks et al., 2004; Reschly & Christenson, 2012). We were also able to successfully fold in social engagement as a fourth component, as recommended recently by literacy scholars (Alexander, 2018; Ivey & Johnston, 2013, 2015). We plan to expand implementation of the WES in different contexts to see if the factor analysis outcomes remain as presented in the current report or if they shift with more data. We will also scrutinise the social engagement items to see if we could further improve the wording to more accurately represent the phenomenon.

## ***Teacher friendly***

Teachers were included in the research and tool design of the WES from the beginning, contributing validity evidence based on scale content, through a focus group with two elementary teachers. As a result of this feedback, the WES was designed with teachers' concerns at the forefront. Accordingly, and in keeping with Moffett's (1992) views on the need for assessment of naturally occurring classroom writing tasks, the WES demonstrates an ease of use for the purposes of formative assessment of young writers that minimises the burden on teachers and positions them to utilise the results in immediately actionable ways. Further refinement of the WES based on teacher feedback (and research results) will contribute to the refinement of an assessment tool that is teacher friendly.

## ***Implications for practice***

The WES is a classroom-friendly formative assessment tool. It includes only 16 items and one open-ended question. For easy administration, we distributed it through Google Forms. Therefore, it is easy and fast to implement – it does not take much valuable classroom time away (~5 minutes) – and it is easy and fast to analyse the results. These attributes, along with its acceptable psychometric properties, make it useful for teachers who are seeking to better inform their writing instruction. Additionally, it is a positive classroom tool because it assesses an often overlooked, yet vitally important, component of writing. When students are engaged in writing they find it enjoyable, they are committed to the task, and they are engaged in strategic metacognitive thought – all of which enhances writing abilities. Similarly, students are more likely to spend time and cognitive energy on writing if they find it engaging, and more time spent writing is likely to support writing growth. Writing can be – indeed, should be – an enjoyable, cathartic, and rewarding activity. The more that researchers and teachers can move students towards that understanding of writing, the better it is for them and for their educational pursuits.

## ***Implications for research***

### ***Validity***

With the WES, we saw a theoretically appropriate inclusion of affective, behavioural, cognitive, and social indicators of writing engagement, which supports previous research conceptualising engagement through these constructs (Alexander, 2018). This measure

should be further analysed in future research, especially in terms of the weaker evidence of validity, which were reflected in factor loadings (especially for social engagement) and a lack of difference in writing engagement based on grade level. However, there were multiple pieces of evidence that supported the validity of the WES as a measure of student writing engagement, such as strong global fit indices, clear discriminant validity indicating a multidimensional factor structure, acceptable reliability of three out of four subscales, and the presence of differences based on gender. Moreover, teacher perceptions of the value of the WES for formative assessment purposes provide an additional layer of validity, in that psychometric validation alone would not justify the widespread use of such a scale if teachers found no value in its implementation for classroom instruction. We recommend future researchers utilise this measure while continuing to investigate multiple types of evidence for its validity.

### **Formative**

While the research base linking self to formative assessment is strong, and has grown in particular in the past 20 years, its application to writers in the early grades for the purposes of writing assessment requires further investigation as well as replication of what existing studies have explored this topic and yielded positive, valid results. Further still, as models of writing engagement are by and large not well developed, leveraging assessments to investigate the factors that affect students' engagement in writing requires the field's serious, renewed attention.

### **Engagement**

Future research on the WES should first expand its implementation. Studying validity and reliability across multiple contexts both across the United States and internationally would allow for a more robust and trustworthy understanding of its validity. Likewise, more in-depth and extended analysis of teachers' reflections on the use of the WES as a formative assessment tool would provide deeper insight into how it can be used and refined. We designed the WES, above all, to support teachers in providing an expanded view of their students' writing – in an easy-to-use, but insightful, tool. Learning more about the degree to which the WES embodies that goal would direct us in refining or promoting its use. A final and significant study that is needed is to analyse WES results in light of students' writing performance. That is, is writing engagement, as measured by the WES, associated with writing performance, as measured by a valid and reliable evaluation of students' writing proficiency. A logical hypothesis is that students who are more engaged in writing would perform better at writing. However, we are unaware of any empirical research that provides such evidence. This is a necessary next step.

### **Limitations**

Our greatest limitations are found in our sample size. Although it is possible to conduct CFAs with lower sample sizes, our analysis would have been much stronger with more participating students. Because of our low sample size, we could not evaluate measurement invariance. Another limitation is that our sample skewed heavily towards the upper

elementary grades in that more than 74% of respondents were in fourth or fifth grade. In the future we would like a more balanced sample in terms of grade level. An additional limitation is our exclusion of items due to poor factor loadings or high residuals. Although our model was stronger when we excluded four items, we would prefer to have a balanced scale and would especially prefer to have more than two items for the behavioural engagement scale.

### ***Future directions***

As mentioned in our implications for research and limitations, future research needs to continue to investigate the validity of the WES. Specifically, future researchers should collect evidence from a larger sample size and with a sample more representative of different grade levels. Because we did find a clear factor structure, we recommend future research uses additional CFAs to see if our factor structure is replicable. Researchers who have a larger sample size could also use CFA to allow inspection of measurement invariance, which would inform our abilities to accurately compare students based on characteristics like gender and grade level. To gather further evidence of validity, future research should collect additional information from students, such as writing performance, to see if relationships between related concepts, such as writing engagement and writing performance, are consistent with theoretical understandings. Although we attempted to do this, our small sample of writing performance prevented any generalisable findings about this relationship.

Further research into the correlation between student self-assessment results and the quality of the texts they are rating would shed further light on the predictive value of the WES. If students' self-assessments can in any way predict or presage the quality of the texts they produce, the value and import of student self-assessment tools like the WES will increase for teachers and students alike.

### **Conclusion**

There is still a great deal to be done in utilising self-assessment for the purposes of formative assessment – particularly for the purposes of devising valid formative assessment tools to assess student writing in the early grades. The findings of this study provide suggestions for the design of formative assessment tools that encourage the behaviours that undergird positive student engagement in writing, which we briefly describe below.

First, formative assessment should incorporate not only evaluation of written products but also assessments of students' overall relationship to writing. The best source of information for both types of assessments come from the students themselves; thus, self-assessment for formative assessment is a powerful tool in teachers' repertoire of instruction. However, this information needs to be provided to teachers in ways which make it easy to analyse at the individual and whole-class levels allowing them to adjust instruction in meaningful and impactful ways for their students as a natural part of their pedagogy and curriculum.

Second, if students are more engaged with writing, generally and at the task level, they are more likely to learn than students who are not (Rogers, 2008). This is important because our goals for students extend beyond writing proficiency towards a lifetime of participation in literate activity. Self-assessment provides an opportunity to both inform

teachers' practice and help researchers understand the multiple dimensions of students' engagement in writing. Standardised tests and large scale international, national, and state level assessments provide comparative benchmarks. However, the unintended consequences of the overemphasis on these kinds of assessments obscures students' role in their own learning and distorts our conceptualisation of writing proficiency. We must account for students in the assessment process, for as Moffett (1992) noted, 'as soon as others want the results of learning more than the learner, the game is over' (p. 27).

## Disclosure statement

No potential conflict of interest was reported by the author(s).

## Notes on contributors

**Paul M. Rogers** is an Associate Professor of Writing Studies at the University of California, Santa Barbara. His areas of interest are writing development, writing instruction, writing assessment, writing research methodologies, and the professional development of teachers.

**Jonathan M. Marine** is a writing instructor at George Mason University where he is also a PhD student in Writing and Rhetoric. His research interests include content analysis, writing engagement, longitudinal writing development, the pedagogy and theory of James Moffett, and the rhetoric of material/public inscription.

**Samantha T. Ives** is a PhD candidate in Educational Psychology at George Mason University. Her research interests include engagement, motivation, and survey validation.

**Seth A. Parsons** is a Professor in the School of Education and the Sturtevant Center for Literacy at George Mason University. He teaches in the Elementary Education and Literacy program areas.

**Ashlee Horton** is a K-5 Instructional Specialist with over twenty years of teaching experience in a variety of settings.

**Chase Young** is a professor in the School of Teaching and Learning at Sam Houston State University. Previously, he taught elementary school and served as a literacy coach.

## References

- Alexander, P. A. (2018). Engagement and literacy: Reading between the lines. *Journal of Research in Reading*, 41(4), 732–739. <https://doi.org/10.1111/1467-9817.12262>
- Amrein-Beardsley, A., & Berliner, D. C. (2003). Re-analysis of NAEP math and reading scores in states with and without high-stakes tests: Response to Rosenshine. *Education Policy Analysis Archives*, 11. <https://doi.org/10.14507/epaa.v11n25.2003>
- Andrade, H. L., Du, Y., & Mycek, K. (2010). Rubric-referenced self-assessment and middle school students' writing. *Assessment in Education: Principles, Policy & Practice*, 17(2), 199–214. <https://doi.org/10.1080/09695941003696172>
- Applebee, A. N., & Langer, J. A. (2006). *The state of writing instruction in America's schools: What existing data tell us*. Center on English Learning and Achievement.
- Baird, J. A., Andrich, D., Hopfenbeck, T. N., & Stobart, G. (2017). Assessment and learning: Fields apart? *Assessment in Education: Principles, Policy & Practice*, 24(3), 317–350. <https://doi.org/10.1080/0969594X.2017.1319337>
- Behizadeh, N. (2014). Adolescent Perspectives on Authentic Writing Instruction. *Journal of Language and Literacy Education*, 10(1), 27–44.

- Berninger, V. W., & Hidi, S. (2007). Mark Twain's writers' workshop: A nature–nurture perspective for motivating students with learning disabilities to compose. In S. Hidi & P. Boscolo (Eds.), *Writing and motivation* (pp. 163–182). Leiden, NL: Brill.
- Berry, R. (2011). Assessment reforms around the world. In B. Adamson, & R. Berry (Eds.), *Assessment reform in education* (pp. 89–102). Dordrecht: Springer.
- Black, P., & Wiliam, D. (1998). Assessment and classroom learning. *Assessment in Education: Principles, Policy & Practice*, 5(1), 7–74 <https://doi.org/10.1080/0969594X.2018.1441807>
- Black, P., & Wiliam, D. (2018). Classroom assessment and pedagogy. *Assessment in education: Principles, Policy & Practice*, 25(6), 551–575.
- Boscolo, P., & Hidi, S. (2007) The multiple meanings of motivation to write. In Rijlaarsdam, G., P. Boscolo, & S. Hidi (Eds.), *Writing and Motivation* (Vol. 19, pp. 1–14)Studies in Writing. Oxford: Elsevier.
- Brookhart, S. M. (2016). Section discussion: Building assessments that work in classrooms. *Handbook of Human and Social Conditions in Assessment*, 351–365.
- Brown, T. A. (2015). *Confirmatory factor analysis for applied research* (2nd ed.). The Guilford Press.
- Brown, G. T. L., & Harris, L. R. (2013). Student self-assessment. In J. H. McMillan (Ed.), *The sage handbook of research on classroom assessment* (pp. 367–393). Sage.
- Brown, G., & Harris, L. R. (2014). The future of self-assessment in classroom practice: Reframing self-assessment as a core competency. *Frontline Learning Research: An Official Journal of EARLI*, 2(1), 22–30. <https://doi.org/10.14786/flr.v2i1.24>
- Calfee, R., Wilson, K. M., Flannery, B., & Kapinus, B. (2014). Formative assessment for common core literacy standards. *Teachers College Record*, 116(11), 1–32. <https://doi.org/10.1177/016146811411601106>
- Camacho, A., Alves, R. A., & Boscolo, P. (2021). Writing motivation in school: A systematic review of empirical research in the early twenty-first century. *Educational Psychology Review*, 33(1), 213–247. <https://doi.org/10.1007/s10648-020-09530-4>
- Cho, M. (2019). The effects of prompts on L2 writing performance and engagement. *Foreign Language Annals*, 52(3), 576–594. <https://doi.org/10.1111/flan.12411>
- Christenson, S. L., Reschly, A. L., & Wylie, C. (Eds.), (2012). *Handbook of research on student engagement*. Springer Science & Business Media.
- Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Erlbaum.
- Dunn, T., Baguley, T., & Brunsdon, V. (2014). From alpha to omega: A practical solution to the pervasive problem of internal consistency estimation. *British Journal of Psychology*, 105(3), 399–412. <https://doi.org/10.1111/bjop.12046>
- Farizka, N. M., & Cahyono, B. Y. (2021). Faculty members' strategies to foster students' learning engagement in writing class. *Journal on English as a Foreign Language*, 11(1), 175–196. <https://doi.org/10.23971/jefl.v11i1.2478>
- Finn, J. D., & Zimmer, K. S. (2012). Student engagement: What is it? Why does it matter? In S. L. Christenson (Ed.), *Handbook of research on student engagement* (pp. 97–131). Springer.
- Fredricks, J. A., Blumenfeld, P. C., & Paris, A. H. (2004). School engagement: Potential of the concept, state of the evidence. *Review of Educational Research*, 74(1), 59–109. <https://doi.org/10.3102/00346543074001059>
- Fredricks, J. A., & McColskey, W. (2012). The measurement of student engagement: A comparative analysis of various methods and student self-report instruments. In S. L. Christenson, A. L. Reschly, & C. Wylie (Eds.), *Handbook of research on student engagement* (pp. 763–782). Springer.
- Graham, S. (2019). Changing how writing is taught. *Review of Research in Education*, 43(1), 277–303. <https://doi.org/10.3102/0091732X18821125>
- Graham, S., & Harris, K. R. (2013). Designing an effective writing program. In S. Graham, C. MacArthur, & J. Fitzgerald (Eds.), *Best Practices in Writing Instruction* (Vol. 2, pp. 3–25).New York, NY: Guilford.

- Graham, S., Harris, K. R., Kiuahara, S. A., & Fishman, E. J. (2017). The relationship among strategic writing behavior, writing motivation, and writing performance with young, developing writers. *The Elementary School Journal*, 118(1), 82–104. <https://doi.org/10.1086/693009>
- Graham, S., & Rijlaarsdam, G. (2016). Writing education around the globe: Introduction and call for a new global analysis. *Reading and Writing*, 29(5), 781–792. <https://doi.org/10.1007/s11145-016-9640-1>
- Guthrie, J. T., Wigfield, A., & You, W. (2012). Instructional contexts for engagement and achievement in reading. In S. L. Christenson, A. L. Reschly, & C. Wylie (Eds.), *Handbook of research on student engagement* (pp. 601–634). Springer US.
- Harris, L. (2011). Secondary teachers' conceptions of student engagement: Engagement in learning or in schooling? *Teaching and Teacher Education*, 27(2), 376–386. <https://doi.org/10.1016/j.tate.2010.09.006>
- Hawthorne, S. (2008). Students' beliefs about barriers to engagement with writing in secondary school English: A focus group study. *Australian Journal of Language and Literacy*, 31(1), 30–42. <https://search.informit.org/doi/epdf/10.3316/ielapa.945771761871219>
- Hayes, A. F., & Coutts, J. J. (2020). Use omega rather than cronbach's alpha for estimating reliability. *Communication Methods and Measures*, 14(1), 1–24. <https://doi.org/10/ggwd6m>
- Hillocks, G. (2003). Fighting back: Assessing the assessments. *The English Journal*, 92(4), 63–70. <https://doi.org/10.2307/3650459>
- Hu, L., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling: A Multidisciplinary Journal*, 6(1), 1–55. <https://doi.org/10.1080/10705519909540118>
- Huot, B. (1990). The literature of direct writing assessment: Major concerns and prevailing trends. *Review of Educational Research*, 60(2), 237–263. <https://doi.org/10.3102/00346543060002237>
- Ivey, G., & Johnston, P. H. (2013). Engagement with young adult literature: Outcomes and processes. *Reading Research Quarterly*, 48(3), 255–275. <https://doi.org/10.1002/rrq.46>
- Ivey, G., & Johnston, P. H. (2015). Engaged reading as a collaborative transformative practice. *Journal of Literacy Research*, 47(3), 297–327. <https://doi.org/10.1177/1086296X15619731>
- Jorgensen, T. D., Pornprasertmanit, S., Schoemann, A. M., & Rosseel, Y. (2021). *semTools: Useful tools for structural equation modeling. R package version 0.5-5*. <https://CRAN.R-project.org/package=semTools>
- Kitsantas, A., & Zimmerman, B. J. (2006). Enhancing self-regulation of practice: The influence of graphing and self-evaluative standards. *Metacognition and Learning*, 1(3), 201–212. <https://doi.org/10.1007/s11409-006-9000-7>
- Kline, R. B. (2016). *Principles and practice of structural equation modeling* (4th ed.). The Guilford Press.
- Kostons, D., Van Gog, T., & Paas, F. (2012). Training self-assessment and task-selection skills: A cognitive approach to improving self-regulated learning. *Learning and Instruction*, 22(2), 121–132. <https://doi.org/10.1016/j.learninstruc.2011.08.004>
- Leahy, S., & Wiliam, D. (2012). From teachers to schools: Scaling up professional development for formative assessment. In J. Gardner (Ed.), *Assessment and Learning* (Vol. 2, pp. 49–71). London: Sage.
- Lee, Y., Jang, B. G., & Conradi Smith, K. (2021). A systematic review of reading engagement research: What do we mean, what do we know, and where do we need to go? *Reading Psychology*, 42(5), 540–576. <https://doi.org/10.1080/02702711.2021.1888359>
- Liu, M., Liu, L., & Liu, L. (2018). Group awareness increases student engagement in online collaborative writing. *The Internet and Higher Education*, 38, 1–8. <https://doi.org/10.1016/j.iheduc.2018.04.001>
- McNeish, D. (2018). Thanks coefficient alpha, we'll take it from here. *Psychological Methods*, 23(3), 412–433. <https://doi.org/10.1037/met0000144>
- Moffett, J. (1992). *Detecting Growth in Language*. Boynton/Cook Pub.
- National Academies of Sciences, Engineering, and Medicine. (2018). *How people learn II: Learners, contexts, and cultures*. The National Academies Press.

- National Writing Project. (2010). *The Analytic Writing Continuum: A comprehensive writing assessment system*.
- Nicol, D. J., & Macfarlane-Dick, D. (2006). Formative assessment and self-regulated learning: A model and seven principles of good feedback practice. *Studies in Higher Education*, 31(2), 199–218. <https://doi.org/10.1080/03075070600572090>
- Olina, Z., & Sullivan, H. J. (2004). Student self-evaluation, teacher evaluation, and learner performance. *Educational Technology Research and Development*, 52(3), 5–22. <https://doi.org/10.1007/BF02504672>
- Panadero, E., Brown, G. T., & Strijbos, J. W. (2016). The future of student self-assessment: A review of known unknowns and potential directions. *Educational Psychology Review*, 28(4), 803–830. <https://doi.org/10.1007/s10648-015-9350-2>
- Parsons, S. A., Malloy, J. A., Parsons, A. W., & Burrowbridge, S. C. (2015). Students' engagement in literacy tasks. *The Reading Teacher*, 69(2), 223–231. <https://doi.org/10.1002/trtr.1378>
- Parsons, S. A., Malloy, J. A., Parsons, A. W., Peters-Burton, E., & Burrowbridge, S. C. (2018). Sixth-grade students' engagement in academic tasks. *The Journal of Educational Research*, 111(2), 232–245. <https://doi.org/10.1080/00220671.2016.1246408>
- Pino-James, N., & Nicolás Pino-James. (2018). Evaluation of a pedagogical model for student engagement in learning activities. *Educational Action Research*, 26(3), 456–479. <https://doi.org/10.1080/09650792.2017.1354771>
- Ramdass, D., & Zimmerman, B. J. (2008). Effects of self-correction strategy training on middle school students' self-efficacy, self-evaluation, and mathematics division learning. *Journal of Advanced Academics*, 20(1), 18–41. <https://doi.org/10.4219/jaa-2008-869>
- Raykov, T. (2001). Estimation of congeneric scale reliability using covariance structure analysis with nonlinear constraints. *British Journal of Mathematical and Statistical Psychology*, 54(2), 315–323. <https://doi.org/10.1348/000711001159582>
- Raykov, T., & Marcoulides, G. A. (2019). Thanks coefficient alpha, we still need you! *Educational and Psychological Measurement*, 79(1), 200–210. <https://doi.org/10.1177/0013164417725127>
- Regan, K., Evmenova, A. S., Boykin, A., Sacco, D., Good, K., Ahn, S. Y., MacVittie, N., & Hughes, M. D. (2017). Supporting struggling writers with class-wide teacher implementation of a computer-based graphic organizer. *Reading & Writing Quarterly*, 33(5), 428–448. <https://doi.org/10.1080/10573569.2016.1221781>
- Reschly, A. L., & Christenson, S. L. (2012). Jingle, jangle, and conceptual haziness: Evolution and future directions of the engagement construct. In S. L. Christenson, A. L. Reschly, & C. Wylie (Eds.), *Handbook of research on student engagement* (pp. 1–19). Springer.
- Rogers, P. M. (2008). *The development of writers and writing abilities: A longitudinal study across and beyond the college-span*. University of California.
- Rosseel, Y. (2012). lavaan: An R package for structural equation modeling. *Journal of Statistical Software*, 48(2), 1–36. <https://doi.org/10/f3r4v8>
- Satorra, A., & Bentler, P. M. (1994). Corrections to test statistics and standard errors in covariance structure analysis. In A. vonEye & C. C. Clogg (Eds.), *Latent variables analysis: Applications for developmental research* (pp. 399–419). Sage.
- Satorra, A., & Bentler, P. M. (2001). A scaled difference chi-square test statistic for moment structure analysis. *Psychometrika*, 66(4), 507–514. <https://doi.org/10.1007/BF02296192>
- Sattler, J. M., & Hoge, R. D. (2006). *Assessment of children: Behavioral, social, and clinical foundations* (5th ed.). J.M. Sattler.
- Savalei, V., & Reise, S. P. (2019). Don't forget the model in your model-based reliability coefficients: A reply to McNeish. *Collabra: Psychology*, 5(1), 36. <https://doi.org/10/gn377t>
- Schildkamp, K., van der Kleij, F. M., Heitink, M. C., Kippers, W. B., & Veldkamp, B. P. (2020). Formative assessment: A systematic review of critical teacher prerequisites for classroom practice. *International Journal of Educational Research*, 103, 101602. <https://doi.org/10.1016/j.ijer.2020.101602>
- Shernoff, D. J. (2013). *Optimal learning environments to promote student engagement*. Springer Science + Business Media.



- Skinner, E. A., Kindermann, T. A., & Furrer, C. J. (2009). A motivational perspective on engagement and disaffection: Conceptualization and assessment of children's behavioral and emotional participation in academic activities in the classroom. *Educational and Psychological Measurement*, 69(3), 493–525. <https://doi.org/10.1177/0013164408323233>
- Skinner, E. A., & Pitzer, J. R. (2012). Developmental dynamics of student engagement, coping, and everyday resilience. In S. L. Christenson, A. L. Reschly, & C. Wylie (Eds.), *Handbook of research on student engagement* (pp. 21–44). Springer.
- Swaffield, S. (2011). Getting to the heart of authentic assessment for learning. *Assessment in Education: Principles, Policy & Practice*, 18(4), 433–449. <https://doi.org/10.1080/0969594X.2011.582838>
- Tabachnick, B. G., Fidell, L. S., & Ullman, J. B. (2019). *Using multivariate statistics* (7th ed.). Pearson.
- Tan, K. H. K. (2001). . *Unpublished PhD Thesis Confirmation Paper*. University of Technology, Sydney.
- Tan, K. H. K. (2012). *Student self-assessment: Assessment, learning, and empowerment*. Singapore: Research Publishing.
- Taras, M. (2010). Student self-assessment: Processes and consequences. *Teaching in Higher Education*, 15(2), 199–209. <https://doi.org/10.1080/13562511003620027>
- Tsao, J. J. (2021). Effects of EFL learners' L2 writing self-efficacy on engagement with written corrective feedback. *The Asia-Pacific Education Researcher*, 30(6), 1–10.
- Upadhyaya, K., & Salmela-Aro, K. (2013). Development of school engagement in association with academic success and well-being in varying social contexts. *European Psychologist*, 18(2), 136–147. <https://doi.org/10.1027/1016-9040/a000143>
- Vygotsky, L. (1978). Interaction between learning and development. In M. Cole, V. John-Steiner, & S. Scribner (Eds.), *Mind in Society* (pp. 79–91). Cambridge, MA: Harvard University Press.
- Walpole, S., Strong, J. Z., & Vitale, M. N. (2021). Instrument development. In N. K. Duke & M. H. Mallette (Eds.), *Literacy research methodologies* (3rd ed., pp. 211–233). Guilford.
- Wang, Z., Bergin, C., & Bergin, D. A. (2014). Measuring engagement in fourth to twelfth grade classrooms: The classroom engagement inventory. *School Psychology Quarterly*, 29(4), 517–535. <https://doi.org/10.1037/spq0000050>
- Weeden, P., Winter, J., & Broadfoot, P. (2002). *Assessment*. Psychology Press.
- William, D. (2006). Formative assessment: Getting the focus right. *Educational Assessment*, 11(3/4), 283–289. [https://doi.org/10.1207/s15326977ea1103&4\\_7](https://doi.org/10.1207/s15326977ea1103&4_7)
- William, D. (2017). Assessment and learning: Some reflections. *Assessment in Education: Principles, Policy & Practice*, 24(3), 394–403.
- Wolsey, T. D., & Grisham, D. L. (2007). Adolescents and the new literacies: Writing engagement. *Action in Teacher Education*, 29(2), 29–38. <https://doi.org/10.1080/01626620.2007.10463446>
- Yancey, K. B. (1999). Looking back as we look forward: Historicizing writing assessment. *College Composition and Communication*, 50(3), 483–503. <https://doi.org/10.2307/358862>