

## Comparing the Construct of “Reading Proficiency” Across Five Commonly Used Reading Assessments: Implications for Policy and Practice

Kristin M. Gehsmann, Ed.D.<sup>1</sup>

Alexandra N. Spichtig, Ph.D.<sup>2</sup>

Jeffrey P. Pascoe, Ph.D.<sup>2</sup>

John D. Ferrara, M. Ed.<sup>2</sup>

<sup>1</sup> St. Michael's College, Colchester, VT USA

<sup>2</sup> Taylor Communications, Inc., Winooski, VT USA

**Conference Paper**  
**Literacy Research Association**  
**December 2017**

## Background

In the US, there is a long history of efforts to synthesize the results of reading research into coherent guidelines and grade-specific standards to help educators plan instruction. Many policy initiatives have emerged from such efforts including No Child Left Behind and the Common Core State Standards for English Language Arts (CCSS ELA; NGA/CCSSO, 2010). (For reviews, see Pearson, 2013; Pearson & Goodin, 2010; Pearson & Hiebert, 2013.)

Critical to the evaluation of these policies are reliable measures of students' literacy achievement. With this in mind, identifying assessments that are valid and reliable, cost-effective, and minimally time-consuming to administer and score is one of the more important matters literacy educators and scholars must consider.

Vast sums of money and weeks of instructional time are spent on testing students every year. Historically speaking, however, assessment data typically have not been utilized effectively (Madaus & Russell 2010/2011; Mertler & Zachel, 2006; Valencia & Wixson, 2000). Further, too often the intended use of assessment data does not sufficiently guide the choice of assessment, or the intended use of assessment data is misaligned with the original purpose for which the assessment was designed. For example, while a standards-based assessment may serve appropriately as an accountability measure to determine how many students within a given school or district met performance standards, the results will not provide educators with specific information to explain why a particular student might be struggling, nor will they help to inform instructional decisions at the student level (Invernizzi, Landrum, Howell, & Warley, 2005; Nichols & Berliner, 2007; Valencia & Buly, 2004; Valencia, Wixson, & Pearson, 2011). Moreover, in some cases, there may be long time gaps between the time an assessment is administered and when results become available, or the results may never become available at all.

## Rationale

The increasing pressure put on schools by standards and accountability systems assures there will be no shortage of available assessments. Helping educators and administrators navigate their many assessment choices is an important issue that deserves attention. While the purpose for assessing students plays a key role in the assessment selection process, logistical considerations may also need to be an important consideration. Thus, the purpose of this study was to compare assessment results from five different instruments that require differing amounts of time to administer. These five assessments were developed across multiple decades, each adhering to the philosophies and standards popular at the time they were created.

Two primary questions guided this inquiry:

- How strong are the correlations between the proficiency scores established by each of the different measures?
- Are correlation coefficients affected by the different methods of assessing proficiency?

After the initial analysis of the relationship between the five literacy measures, we added the following post-hoc question concerning the relationship between reading proficiency and achievement in math: Is there a relationship between the proficiency scores on reading tasks and math assessments?

## Methods

### Participants

A total of 249 students were included in this study (141 females and 108 males. Thirty-four percent of the students received free or reduced price lunch. English language learners (ELLs) comprised 12% of the sample 15% were on IEPs . The sample was 73% White, 10% African-American, 8% Asian, 8% multi-racial, and 1% other. Students identified as Hispanic/Latino ethnicity comprised 0.5 % of the sample.

### Measures

Literacy proficiency data were collected between April and May of the 2015-16 school year using five different assessment instruments:

1. Reading proficiency was assessed using the Group Reading Assessment Diagnostic Evaluation (GRADE), a timed, norm-referenced paper and pencil diagnostic reading measure (Williams, 2001). The GRADE takes 90 minutes to administer and includes two parallel test forms for each grade level (Forms A and B). Each grade level test is divided into subtests to evaluate core literacy skills. The fourth- and fifth-grade test includes (a) listening comprehension, (b) sentence comprehension, (c) passage comprehension, and (d) vocabulary. The test calculates standardized subtest scores, comprehension composite scores that include sentence and passage comprehension, and total test scores that combine comprehension and vocabulary scores.
2. Reading proficiency was also assessed using InSight, a 30-minute adaptive, web-based, silent reading screener (Taylor Associates Communications, Winooski, VT USA). InSight provides measures of reading comprehension, vocabulary, reading rate, and motivation, as well as a composite reading proficiency score. Students receive no instructional support while completing InSight. As such, this assessment gauges reading proficiency in a format more akin to authentic reading tasks and yields results that can readily be compared with nationally normed standardized test results (Reading Plus, 2017).
3. The Elementary Spelling Inventory was used to determine students' word knowledge and stage of literacy and spelling development (Bear, Invernizzi, Templeton, & Johnston, 2016). The Center for Research in Education Policy has found this inventory to be a valid and reliable tool for this purpose (Sterbinsky, 2007). Unlike traditional spelling tests, developmental spelling inventories assess students' word knowledge by evaluating the orthographic features students use when spelling twenty-five words specifically chosen for the orthographic features they contain. Students are awarded points for using each of these features as well as for spelling words correctly. The total score combines feature and word points. The power score represents

the number of words spelled correctly. A student's stage of development can be determined by analyzing their feature points and/or power score (Bear et al., 2016).

4. Reading efficiency was evaluated using a portable eye-movement recording system that measures students' reading rate and visual activity while reading nationally normed 100-word passages with demonstrated comprehension (Visagraph; Taylor Associates, 2009; see Spichtig et. al., 2016). Typically, a recording takes between 5-10 minutes to complete. The assessment provides data on a number of measures including silent reading rate, eye fixations, fixation durations, and regressions. For the purpose of this study, only the measure of silent reading rate is used.

5. English Language Arts (ELA) and Math scores on the Smarter Balanced assessment (Smarter Balanced Assessment Consortium, 2016) were used as a measure of students' achievement of the Common Cores State Standards (CCSS). Although Smarter Balanced assessments (SBAC) are untimed, estimates of three to four and one half hours per subject area are provided. The SBAC was used in at least fifteen states as part of their statewide accountability programs in 2015-2016 (EdWeek, 2016).

## Results

Outlined in Table 1 are the correlation coefficients (Pearson's  $r$ ) calculated using scores generated by each of the assessments. The results highlight a significant overlap across all assessment measures, even including the SBAC Math measure. All correlations were significant ( $p < .001$ ), including correlations with the motivation measures (self-efficacy and interest). The strongest correlations were found among the overall reading proficiency measures (GRADE Standard Score, SBAC ELA Scale Score, and InSight Proficiency Index). Among these three measures, the lowest correlation coefficient was found between the GRADE Total Standard Score and SBAC ELA Scale Score ( $r = .797$ ). The strongest correlation ( $r = .839$ ) was detected between the GRADE Standard Score and the InSight Proficiency Index, although the correlation between the InSight Proficiency Index and the SBAC ELA Scale Score was only slightly less strong ( $r = .834$ ). Surprisingly, even scores on the SBAC Math assessment were strongly correlated with all three reading proficiency measures. The highest correlation was found between the SBAC ELA Scale Score and the SBAC Math Scale Score ( $r = .807$ ).

Reading rate measures and spelling inventory scores were more strongly correlated with the three reading proficiency measures than with each other. This may not be surprising considering the contribution of the assessed skills to overall reading proficiency. It is also important to note that many students topped out the Elementary Spelling inventory and thus limited the range of information the inventory could have provided.

Table 1. Pearson Correlation Coefficients Across Multiple Assessments

		GRADE	SBAC		InSight			Visa	Spelling			
		Total SS	ELA SS	Math SS	Prof. Index	Rate	Motivation SE I	Eff.	PS	FP	TP	
GRADE	Total SS	1	.797*	.689*	.839*	.629*	.429*	.454*	.590*	.540*	.568*	.574*
	SBAC	ELA Sc. Score	.797*	1	.807*	.834*	.619*	.420*	.409*	.588*	.636*	.655*
Math Sc. Score		.689*	.807*	1	.737*	.492*	.299*	.275*	.491*	.573*	.586*	.589*
InSight	Prof. Index	.839*	.834*	.737*	1	.691*	.437*	.442*	.651*	.598*	.627*	.632*
	Reading Rate	.629*	.619*	.492*	.691*	1	.315*	.347*	.652*	.446*	.462*	.464*
	Self-Eff. (SE)	.429*	.420*	.299*	.437*	.315*	1	.701*	.460*	.424*	.435*	.443*
	Interest (I)	.454*	.409*	.275*	.442*	.347*	.701*	1	.421*	.328*	.334*	.339*
Visa	L4 Rate	.590*	.588*	.491*	.651*	.652*	.460*	.421*	1	.449*	.480*	.480*
Spelling	Power Stage	.540*	.636*	.573*	.598*	.446*	.424*	.328*	.449*	1	.948*	.960*
	Feature Points	.568*	.655*	.586*	.627*	.462*	.435*	.334*	.480*	.948*	1	.997*
	Total Points	.574*	.662*	.589*	.632*	.464*	.443*	.339*	.480*	.960*	.997*	1

\*. Correlation is significant at <.001 (2-tailed).

### Implications

This study suggests a significant overlap in reading proficiency scores across assessments regardless of the primary focus of the assessment, the time it takes to administer the assessment, or time period during which a given assessment was developed. Some underlying common factors that contribute to reading proficiency appear to be evaluated regardless of the domain assessed. The high correlation coefficients between the literacy assessments and the SBAC Math assessment raise questions regarding to the extent to which the SBAC Math assessment is indeed an assessment of math and the degree to which students' demonstration of their math skills on this assessment is affected by their level of reading proficiency. An alternative consideration is whether reading proficiency is predictive of math proficiency, or vice versa. This relationship warrants further exploration.

Strong correlations such as those observed in this study may support the value of providing educators with additional guidance concerning how to select the best assessment for their intended purpose(s). Considerations include the time required to administer and score the assessment, the ease with which it is administered, the format of the test, and how useful the data are when planning instruction for individual students.

### Directions for Future Research

As new standards and policy initiatives emerge in the years to come, the assessments used for accountability purposes will likely change as well, and at a considerable cost to taxpayers. Replicating this study with a larger sample of students could potentially influence the assessment decisions states and local education agencies make in future iterations of policy implementation, potentially resulting in considerable savings, both in terms of dollars and human capital.

### References

- Bear, D., Invernizzi, M., Templeton, D., Johnston, F., (2016). *Words Their Way: Word Study For Phonics, Spelling, and Vocabulary*. Boston, MA: Pearson.
- EdWeek (2016, April 1). State testing: An interactive breakdown of 2015-2016 plans. (Retrieved from: <http://www.edweek.org/ew/section/multimedia/state-testing-an-interactive-breakdown-of-2015-16.html>)
- Invernizzi, M., Landrum, T. J., Howell, J.L., Warley, H.P. (2005). Toward peaceful coexistence of test developers, policymakers, and teachers in the era of accountability. *Reading Teacher* 57(7), 610-618.
- Madaus, G. & Russell, M. (2010/2011). Paradoxes of high stakes testing. *Journal of Education*, 190(1/2), 21-30.
- Mertler, C. A., & Zachel, K. (2006). Data-driven instructional decision making: An idea (and practice) whose time has come. *Principal Navigator*, 1 (3), 6–9.
- National Governors Association Center for Best Practices & Council of Chief State School Officers (NGA/CCSSO; 2010). *Common core state standards for English language arts & literacy in history/social studies, science, and technical subjects, Appendix A*. Washington, DC: Authors. Retrieved from Common Core State Standards Initiative website: [http://www.corestandards.org/assets/Appendix\\_A.pdf](http://www.corestandards.org/assets/Appendix_A.pdf)
- Nichols, S. L. & Berliner, D. (2007). *Collateral damage: How high-stakes testing corrupts America's schools*. Cambridge, MA: Harvard Education Press.
- Pearson, P. D. (2013). Research foundations of the Common Core State Standards in English language arts. In S. Neuman and L. Gambrell (Eds.), *Quality reading instruction in the age of Common Core State Standards* (pp. 237-262). Newark, DE: International Reading Association.
- Pearson, P.D., & Goodin, S. (2010). Silent reading pedagogy: A historical perspective. In E.H. Hiebert & D.R. Reutzel (Eds.), *Revisiting silent reading: New directions for teachers and researchers* (pp. 3-23). Newark, DE: International Reading Association.
- Pearson, P. D., & Hiebert, E. H. (2013). National reports in literacy: building a scientific base for practice and policy (pp. 1133-1149). In A.E. Alerman, N.J. Unrau, & R.B. Ruddell (Eds.). *Theoretical models and processes of reading*. Newark, DE: International Reading Association.
- Reading Plus. (2017). *Silent Reading Assessment*. Retrieved from: <https://www-cdn.readingplus.com/wp-content/uploads/2015/03/InSight-Measuring-Three-Domains.pdf>



- Spichtig, A.N., Hiebert, E.H., Vorstius, C., Pascoe, J.P., Pearson, P.D., & Radach, R. (2016). The decline of comprehension-based silent reading efficiency in the United States: A comparison of current data with performance in 1960. *Reading Research Quarterly, 51*(2), 239–259.
- Taylor, S.E. (2009). *Visagraph III: Implementation guide*. Huntington Station, NY: Taylor Associates Communications, Inc.
- Valencia, S. (2011). Using assessment to improve teaching and learning. In S. J. Samuels & A. E. Farstrup (Eds.), *What the research has to say about reading instruction* (4<sup>th</sup> ed.). Newark, DE: International Reading Association.
- Valencia, S. & Buly, M. (2004). Behind test scores: What struggling readers really need. *Reading Teacher, 57*, 520-531.
- Valencia, S. & Wixson, K. (2000). Policy-oriented research on literacy standards and assessment. In M. L. Kamil, P. Mosenthal, P. D. Pearson & R. Barr (Eds.), *Handbook of reading research* (vol. 3, pp. 909-936). Mahwah, NJ: Erlbaum.
- Valencia, S., Wixson, K., & Pearson, P.D. (2011). Issues in developing reading comprehension assessments aligned with the Common Core State Standards. Paper presented at the 61<sup>st</sup> Annual Conference of the Literacy Research Association, Jacksonville, FL.