Carnegie Foundation
for the Advancement of Teaching

RESEARCH BRIEF

Case Study of ORF as Practical Measure Brief

Carnegie Foundation
for the Advancement of Teaching

AUTHORS

Nathaniel A. Dewey
Director of Improvement Science Research
The CoLab at Improving Education

Nat is focused on research into the quantitative methodology of improvement science as well as the development of practical tools to assist in school improvement. Previously, Nat was an assistant research scientist at Johns Hopkins University and a high school science teacher in Connecticut.

Marc L. Stein
Executive Director
The CoLab at Improving Education

Marc is focused on learning about the method of improvement science and how to support organizations to learn from improvement science. Prior to the CoLab at Improving Education, Marc was the managing director of the Baltimore Education Research Consortium (BERC) and was an associate professor of education policy at Johns Hopkins University.

In this post we describe the identification and use of Oral Reading Fluency (ORF), a common and widely used Curriculum Based Measurement (CBM) measure, as a component of the measurement system for the Baltimore Secondary Literacy Improvement Community (BSLIC), a networked improvement community in the Baltimore City Public Schools that is focused on improving the number of students who reach grade-level literacy proficiency standards in middle school. We conclude with a discussion of insights and lessons learned from using ORF as a practical improvement measure.

Curriculum Based Measurement and Improvement

You might be surprised to learn that there is a class of assessments, widely used in districts and classrooms, with many decades of research supporting them, that can be a ready source of practical improvement measures.

Developed by special education researchers beginning in the 1970s, CBM is a method for monitoring student progress toward instructional goals and determining if a student is responding as desired to an intervention. CBM encompasses a problem-solving approach, a framework for instructional decision making, and the development of "general outcome measures" that are simple to use and are valid and reliable for monitoring student progress.

Our insight is that the core elements of CBM are the same ones taken up by improvers in pursuit of understanding the question "Is our change an improvement?"

Reading the abstract to Stanley Deno's seminal article Curriculum-Based Measurement: The Emerging Alternative[1], one cannot help but recognize an immediate connection to developing practical measures for continuous improvement in education. It is worth quoting at length:

> Despite general agreement that we should routinely assess the student performance outcomes of instruction, general agreement regarding how this should be done does not exist. Commercially distributed achievement tests are not always congruent with curriculum objectives and teachers tend not to value the information obtained from them. Informal observation of performance is the approach used and preferred by teachers. Unfortunately, the reliability and validity of teachers' informal observation of student academic performance is unknown...Through standardizing observation of performance in the curriculum CBM generates reliable data that is valid with respect to widely used indicators of achievement such as achievement test scores, age, program placement, and teachers' judgments of competence. These data are now being used to make screening, referral, IEP planning, pupil progress, and program outcome decisions" (p 219).

To our eyes he is talking about practical measures and the science of improvement in 1985!

A critical component of CBM is the set of measures that have been developed to be used to progress monitor students. The properties that make for good CBM measures are properties that are also useful in practical improvement measures: they are sensitive to change in focal skill, they are reliable and valid for the skill of interest, they can be given frequently due to their simplicity and efficiency, they are easily understood, and they can be used for decision making given appropriate frameworks.[2]

Additionally, some CBM measures are also widely used in progress and performance monitoring assessment systems (e.g., DIBELS) used by districts throughout the United States. This means that they are often available for "seeing the system" type analyses at the initiation of improvement projects.

We encourage you to explore the wealth of theoretical and empirical research on CBM.[3] The insights and lessons learned from 50 years of research and practice are of high value to improvers in the field and improvement as a field.

Identification of ORF as a Practical Measure

The decision to use ORF, a CBM measure, as a practical improvement measure emerged over time in the early stages of BSLIC. At this time the hub, teacher fellows, and the measurement and analytics teams were learning rapidly about the working theory of improvement, the operations of improvement, and the measurement system for improvement. This is a very dynamic stage in an improvement project. Over the course of this period in the project there were several strands of work that came together and made clear to everyone that ORF was the best option for a practical improvement measure.

First, and perhaps most importantly, the Literacy Continuous Improvement Coordinator who was leading the project strongly advocated for a focus on fluency as a high-leverage entry point for our improvement work and focus of the working theory of the network.[4] This focusing was critical and emerged from her deep professional and academic expertise in middle grades literacy, engagement with external and academic literacy experts brought in by the Hub, learning directly from teacher fellows doing improvement, and learnings emerging from measurement and analytic work.

Second, a subgroup of hub members that included the Literacy Continuous Improvement Coordinator, the measurement and analytics team, and colleagues at the Carnegie Foundation began searching for practical measures of early literacy skills. Over many months the team explored a wide space of candidate measures. Through weekly discussions we were able to define the needed attributes for our practical measures. Engagement with literacy and measurement experts helped us

to discover new candidate measures and vet our thinking. Small pilot tests with improvement fellows facilitated our learning about the actual (not theoretical) practicalness of candidate measures. Through this process ORF emerged as the strongest candidate in terms of its practicalness and direct connection to the emergent focus on fluency.[5]

Third, the measurement system for improvement was being developed to meet several design criteria, three of which are relevant here. It needed to be able to measure improvement by demonstrating progress. The measurement cadence needed to align with the operational cadence of teacher fellows' improvement work. Further, any measures collected by fellows should feed back useful analyses to fellows and the hub in time to support learning in the operational cadence. As we describe in the Properties of ORF section at the end of this post, ORF has properties that are aligned with these design criteria. This helped to elevate it above other measures under consideration.

Finally, our research on the measurement properties of ORF led us to the insight we shared at the beginning of this article: ORF was developed to be used within a system of methods designed to support the disciplined inquiry of educators. It is worth quoting Deno at length again on this point,

> The CBM procedures were developed as part of a larger program of research directed toward designing a practically feasible and effective formative evaluation system that special education teachers could use to build more effective instructional programs for their students. As part of that formative evaluation system, it was necessary to create a simple, reliable, and valid set of measurement procedures that teachers could use to measure frequently and repeatedly the growth of their students in the basic skills of reading, spelling, and written expression. When these procedures are used within the context of the local school's curriculum, they become CBM (Deno, 1993, p 3).[6]

Beyond the properties of CBM measures, like ORF, that make them practical, it is the system in which they are used that makes them powerful tools for learning and decision making. This is precisely what we are intending to do with continuous improvement practice within a science of improvement in education. If you want to learn more about the ORF measure and its potential as a practical measure, please see the Properties of ORF section at the end of this post.

In future writing we will explore in greater depth the connections between single-subject designs, CBM, and the science of improvement in education. For now we encourage you to explore this literature and hope this discussion has seeded fertile ground for thinking about your own improvement projects.

Lessons Learned Using ORF as a Practical Measure of Improvement

Many of the lessons learned that we are sharing here may seem to be quite obvious. They certainly did to us in retrospect as we were reflecting and writing. At the time, however, these lessons were not yet learned and thus were not so obvious.

A meta-learning, if you will, for us from developing and implementing ORF as a practical measure has been that the disciplined inquiry and cycles of learning that are hallmarks of improvement can be directly applied to the development of practical measures and the measurement system of improvement surrounding them. Practical measures may at some point be able to be used in a general off-the-shelf manner. This is perhaps a worthy goal for us to strive toward, as it may help improvement become more accessible to a wider audience in education. However, we have learned that measures like ORF, with obvious practical attributes, become more practical and thus more powerful for improvement when we turn the improvement sensibility and tools on them and our practice in designing a measurement system. Again, this insight sounds painfully obvious in the writing but assuredly is less so in real-world practice.

While a measure may be practical in theory, that does not make it so in practice.
Candidate practical measures should be subjected to their own improvement learning cycles. ORF was initially pilot-tested with a small number of improvement fellows and a small number of students. Only through safe-to-fail tests under real-world conditions is it possible to determine if a measure is practical in practice and to improve its practicalness. These small tests provided important learnings for fielding ORF at scale. We were able to discover the range of administration times that could be expected in these small tests. We learned that ORF could be collected virtually but was fragile to technological issues. Perhaps most importantly, we learned that there were benefits beyond just data collection of ORF being administered by fellows versus other support staff. Administering ORF gave fellows direct knowledge of each student's fluency by hearing them read aloud, which enriched the fellows' pedagogical and improvement work.

The cadence of practical measurement should match the cadence of learning cycles in the improvement work.
ORF can theoretically be collected on a weekly cadence; however, it would not be practical to do so given the time costs of administering to more than one student. In BSLIC, learning cycles were organized into multi-month action periods that were composed of two-week blocks of fellow improvement work and group reflective huddles. For this operational cadence a monthly collection of ORF provided sufficient data for improvement progress monitoring, presented feedback to fellows, and was practical to collect.

Practical measures are often considered "noisy." This variation is important and should not be ignored.
Attending to variation is a core principle of the science of improvement. In practice, though, it can be a difficult lesson to learn as it is counter to the traditional inclination to see variation as a problem. This is especially true with substantial or unusual variation, which can occur with practical measures. Student ORF performances can vary greatly across a school year. Sometimes a student's reading rate will rise excitingly and other times it will fall unexpectedly. One interpretation of this variation is that the student's true reading rate is being obscured by "noise." In this framing, the noise is seen as a nuisance to be removed if possible or, more likely, modeled away as statistical uncertainty.

When using ORF as a practical measure of improvement, an alternative framing is possible that reverses the traditional inclination to label unexplained variation as noise. Under this interpretation, all variation is considered an important source of information to be explored further. Let's consider an example.

Despite the work of researchers to normalize the difficulty of passages used in ORF, there can be unexpected variation, both positive and negative, that may be due to some passages being easier or harder for some students. Naturally, improvement teams should try to reduce the variation by changing out poor-performing passages over time. Importantly, though, improvement teams should also investigate why some passages are harder or easier for students. Learning from these opportunities can provide valuable new insights on passage elements that are particularly challenging for some students. Digging into this variation could lead to a better understanding of another facet of the difficulties that students might face on reading tasks both on standardized assessments and in authentic situations. This is an important learning that could easily be thrown away as noise.

Endnotes

[1] Deno, S. L. (1985). Curriculum-Based Measurement: The Emerging Alternative. Exceptional Children, 52(3), 219–232. https://doi.org/10.1177/001440298505200303;

[2] See for example: Fuchs, L. S., & Deno, S. L. (1991). Paradigmatic distinctions between instructionally relevant measurement models. *Exceptional Children*, 57(6), 488-500 and Van Norman, E. R., Nelson, P. M., Parker, D. C. (2018). A comparison of nonsense-word fluency and curriculum-based measurement of reading to measure response to phonics instruction. Sch Psychol Q. 33(4):573-581. doi: 10.1037/spq0000237

[3] CBM is a long-running and active area of research. The following review articles and book may be good starting points for the interested reader: Espin, C., McMaster, K. L., Wayman, M. M., & Rose, S. (Eds.) (2012). A measure of success: The influence of curriculum-based measurement on education. U of Minnesota Press; Fuchs, L. S. (2004). The past, present, and future of curriculum-based measurement research. School Psychology Review, 33(2), 188-192; Tindal, G. (2013). Curriculum-based measurement: A brief history of nearly everything from the 1970s to the present. International Scholarly Research Notices, 2013.

[4] We are focusing here only on the fluency aspect of the working theory of improvement that continues to evolve in BSLIC. Throughout, BSLIC has continued work on writing, and fluency has evolved to become one aspect of a focus on the science of literacy in the middle grades.

[5] We investigated a wide range of commercially available and open-source measures that included, among others, the Test of Word Reading Efficiency (TOWRE 2), the Gray Oral Reading Test (GORT-5), various measures of vocabulary and writing, as well as ORF.

[6] Deno, S. (1993). Curriculum based measurement. In J. Kramer (Ed.), *Curriculum-Based Measurement —Complete Work* (pp. 1-24). Buros-Nebraska Series on Measurement and Testing. Accessed from: https://digitalcommons.unl.edu/buroscurriculum/11

[7] Baker, D. L., Biancarosa, G., Park, B. J., Bousselot, T., Smith, J.-L., Baker, S. K., Kame'enui, E. J., Alonzo, J., & Tindal, G. (2015). Validity of CBM measures of oral reading fluency and reading comprehension on high-stakes reading assessments in Grades 7 and 8. *Reading and Writing*, *28*(1), 57–104.

[8] Hasbrouck, J., & Tindal, G. A. (2006). Oral reading fluency norms: A valuable assessment tool for reading teachers. *The Reading Teacher*, *59*(7), 636–644.

[9] Van Norman, E. R., Nelson, P. M., & Parker, D. C. (2018). A comparison of nonsense-word fluency and curriculum-based measurement of reading to measure response to phonics instruction. *School Psychology Quarterly: The Official Journal of the Division of School Psychology*, *American Psychological Association*, *33*(4), 573–581.

Properties of ORF

Name(s):
Oral Reading Fluency (ORF), CBM-R, Passage Reading Fluency (PRF)

Use case:
Measuring early literacy development of middle grades students receiving small-group instruction in phonics and fluency

Task:
The student reads a grade-level passage aloud for 1 minute while an instructor listens and marks any words read incorrectly. At the end of 1 minute, the instructor records how many total words the student read and how many of these words were read incorrectly. The instructor then calculates the words correct per minute (WCPM) as the number of total words read minus the number of words read incorrectly.

Unit of Measure:
Words Correct Per Minute (WCPM)

Frequency Captured:
1 per month (maximum 1 per week)

Expected Change over Time:
In middle grades, growth of ~0.5 WCPM per week (50th percentile growth rate for DIBELS 8th Edition in middle grades)

Predictive Validity:
ORF has been shown in academic literature to be predictive of ELA scores on standardized tests in middle grades[7]. This makes it amenable to use as a practical measure of drivers related to accelerating the acquisition of foundational and early literacy skills and associated aims focused on improving students' ELA performance on state standardized tests.

Responsiveness:
ORF measures reading rate and accuracy, which together with prosody (reading expression) define reading fluency. Unsurprisingly, ORF has been shown to be responsive to reading fluency instruction.[8] Also, for students who need phonics, ORF has been shown to be responsive to phonics instruction as well.[9]

Practicalness:
Administering ORF takes 2-3 minutes per student, which for a middle grades teacher with multiple sections of ELA is a large commitment of their class time. Fortunately, the ORF task has value in and of itself as the teacher gets to hear their students read, which is a valuable pedagogical strategy for early readers.