Maximizing Benchmark Assessments

Increasing Student Achievement and Success

Adisack Nhouyvanisvong, Ph.D.

August 2016

Naiku is a next generation assessment platform, providing teachers with comprehensive assessment tools to help teachers collect data about their students to make informed instruction.

7805 Telegraph Road, Suite 111
Bloomington, MN 55438
Phone: 612-356-2458
E-Mail: info@naiku.net
Web: www.naiku.net

Introduction

The importance of formative assessment—*assessment for learning*—is clear. Teacher's on-going use of assessment to guide and inform instruction has been shown to be one of the most effective instructional strategies to improve student learning (Black & Wiliam, 1998). Summative assessments such as end-of-year annual state tests are also important as they can inform educational policy. However, annual state test results often come too late to impact and inform instruction and student learning.

In between the on-going formative assessments and the once-a-year summative assessments lies benchmark assessments. These assessments can serve the best of both worlds. They can inform teacher instruction and support student learning while also providing school administrators with important information about long-term learning goals that may influence educational policy.

This paper provides a definition of benchmark assessments, the role and purpose of benchmark assessments in a comprehensive assessment system, keys to successful implementation of benchmark assessments and example suggestions.

Definition of Benchmark Assessments

Benchmark assessments are common assessments administered throughout the school year that provide teachers immediate feedback on how students perform relative to short- and long-term learning goals or academic standards. Though they often focus on reading and mathematics skills, benchmark assessments can be used to measure other academic skills. Benchmark assessments are relatively short, taking about an hour per subject, and reflect the priority state academic standards or district learning targets.

Benchmark assessments are typically given periodically throughout the school year, ranging from three times a year to once a month. Administration of benchmark assessments occur at specified times during the academic sequence. Results from benchmark assessments can inform teacher instruction and planning, decision-making at the classroom, school, and district levels, and inform educational policy.

Role in a Comprehensive Assessment System

A balanced and a comprehensive assessment system should provide important data to teachers and administrators to serve classroom, school, and district improvement needs. Benchmark assessment is an integral part of a comprehensive assessment system.

Formative assessment is embedded in classroom instruction. It provides immediate feedback to both teachers and students and allows both to adjust instruction and learning relative to short-term learning goals and targets. Formative assessment data help teachers plan instruction. Annual summative assessments provided administrators and policy makers with information on how students perform relative to annual learning goals or academic standards. Summative end-of-year assessment data help inform educational policy.

Benchmark assessments serve an important role in a comprehensive system. They occupy that important space between frequent formative assessment and end-of-year summative assessment (Data Use, n.d). Unlike formative assessment, they are typically uniform in timing and content across classrooms and schools. Therefore, benchmark assessments provide data that can be aggregated at the classroom, grade, school, and even district levels. They are given more frequently than summative assessments, therefore can provide school and district decision-makers and teachers with interim indication of student learning and progression toward annual goals. The immediate information provided to teachers can also help them plan and adjust subsequent instruction at the classroom level.

Purpose of Benchmark Assessments

Benchmark assessments serve many import purposes as part of the comprehensive assessment system. Four important purposes from Data Use (n.d) are highlighted below:

1. **Communicate expectations for learning.** It is important for students, teachers and parents to know what knowledge and skills are important to learn and how learning will be measured. Benchmark assessments communicate these

expectations frequently to the various stakeholders. It is important vary the benchmark assessment tasks and item types (i.e., include essays, performance tasks, constructed response items, not just multiple-choice items) to communicate that various knowledge and approaches to learning are valued.

2. **Plan instruction**. Benchmark assessments, like formative assessments, provide teachers with immediate and actionable data about student learning and performance. Teachers need information to develop and adjust curriculum and instruction to best serve their students. To do so, benchmark assessment results must be aligned to content and learning targets and provide feedback on students' strengths and weaknesses relative to those learning targets.

3. **Monitor and evaluate learning**. Benchmark assessments can be used to monitor and evaluate learning. They provide educators with information on the effectiveness of educational programs and curricula. Data from benchmark assessments can help administrators make mid-course corrections or modification where student performance is lagging or highlight an educational program or curriculum that helps students excel.

4. **Predict future performance**. With annual statewide accountability tests given at the end of the school year, benchmark assessment results can provide data to predict whether students, classes, schools, or districts are on course to meet those year-end goals. Results can be provided at individual and group levels to identify those who need more help and those who excel and may benefit from more advanced instructional programs.

Quality Criteria for Benchmark Assessments

Many districts purchase benchmark assessments from commercial test publishers. Others develop their own to better meet their needs and match their content standards or learning targets. Regardless of whether they are purchased or newly developed, benchmark assessments, like all assessments, should meet certain criteria for quality. Below, four criteria from Data Use (n.d.) for the selection and use of benchmark assessments are summarized:

1. **Validity**. Validity is often thought of as a criterion of whether an assessment measures what it purports to measure. More accurately, it is the extent to which the inferences made from the assessment results support the purpose for which it is used. A benchmark assessment may be valid for one purpose but have little validity for another. When evaluating benchmark assessments for validity, it is important to consider whether the assessment is assigned with district and school learning goals, provides reliable information for the intended use and score interpretation, is instructionally sensitive, provides useful reporting for intended users and purposes.

2. **Reliability**. Reliability refers to the expected consistency of test scores. Assessments with high measures of reliability provide consistent results across time, across forms, and across settings. When evaluating assessments for reliability, users should examine the item difficulty (appropriate difficulty is in the range of 0.2 to 0.8), item discrimination (desired values are greater than +0.3), and reliability coefficient (test-retest, parallel forms, or internal reliability coefficients should b 0.8 or greater.)

3. **Alignment**. Benchmark assessments should be aligned to the standards or learning targets that the school or district is teaching. Alignment of the assessment to the learning goals is paramount. When evaluating benchmark assessments for alignment, also consider the framework used to develop the assessment and items, the distribution of items by cognitive level demands, and the range of items for diagnosing specific learning strengths and weaknesses.

4. **Utility**. Ultimately, benchmark assessments must have utility. The assessment results must be useful and help teachers and administrators make effective instructional decisions. When considering benchmark assessments for utility, look for assessments that are user-friendly,

easy to administer, scored in a timely fashion, and include robust reports useful for different stakeholders.

Maximizing the Benefit of Benchmark Assessments

The key to maximizing the benefit of benchmark assessments is to optimally find the balance between frequent formative assessment and end-of-year summative assessment. As presented in this paper, benchmark assessments occupy that critical space in between the two and should provide a link between them.

Implemented properly, benchmark assessments consistently provide immediate, valuable, and valid feedback about student performance to teachers to inform instruction, while also providing uniformity so data can be aggregated and compared at higher levels such as the grade, school, or district. This, inherently, is the key to maximizing the benefit – increasing the frequency of benchmark assessments while maintaining the validity.

Some points to consider:

- Are your benchmark assessments given once or twice a year? If so, the benchmark is weighted too far towards summative assessment and there is opportunity to help inform instruction with more frequent assessment.

- Are your benchmark assessments adjustable, such that you can modify the amount and sequence of standards assessed? If so, you can implement benchmarks more frequently and in alignment with curriculum pacing.

- Can staff easily create their own benchmark assessments, drawing from a valid, standards-aligned item bank? If so, this provides significant flexibility to create more frequent, smaller benchmark assessments.

Benchmark Assessments with Naiku

Naiku provides all the tools necessary for schools and districts to implement benchmark assessments.

Included with Naiku is the Inspect® item bank and assessment suite, a user-friendly assessment creation and administration engine, and a scoring and reporting system that provides immediate and useful feedback to students, teachers, and administrations.

Inspect Assessment Suite

Naiku includes over 300 pre-built assessments from Inspect to fully support a school's assessment needs for the entire year. Inspect offers a comprehensive suite of assessments including Interim Assessments, End of Year Summative Assessments, and Performance Tasks. These pre-built assessments are made up of blueprints, diagnostics, and benchmarks for all tested grade levels and includes English language development assessments to benchmark progress in language acquisition (English Language Learner Assessment; ELLA). These included assessments can be used as-is, or customized to align with the sequence and pacing from the school.

The included Inspect Interim Assessments are designed for grades K - 11 for English language arts, and grades K - 8 for math, as well as high school math for both integrated and end - of - course pathways. Each interim assessment is named by color to represent the approximate time of year the assessments should be given.

In order to measure student progress with Standards, the Summative Assessments are included as part of the library of pre-built assessments. The Inspect Summative Assessments feature various innovative item types. Two similar forms of summative assessments are available for grades 3 - 11 for English language arts, and grades 3 - 8 and high school for mathematics with support for both integrated and end-of-course pathways. The two forms allow for districts to analyze student growth on the same standards, skills, and concepts throughout the school year.

Also included are Inspect End-of-Year assessments (EOY) for grades 3-11 for ELA and grades 3-8 and high school (both integrated and end-of-course pathways) for math as part of the library of pre- built assessments. The ELA End-of-Year assessments include a variety of evidence-based selected response items and technology-enhanced constructed response items,

while the mathematic assessments include a variety of innovative item types.

In order to assess students on their ability to utilize the knowledge and skills they have learned, Inspect Performance Tasks in both ELA and math for grades 3-8 and high school are included as part of our library of pre-built assessments. Each Performance Task includes a student booklet and teacher booklet including detailed rubrics for ease of scoring. The Performance Tasks assess students on multiple standards focusing on real - world topics in which they must use their combined knowledge and skills gained.

Inspect Item Bank

Inspect also includes an item bank featuring over 60,000 standards-aligned items. Items are aligned to Common Core State Standards and to specific state standards. The item bank allows schools to develop their own benchmark assessments to match their own learning targets and instructional pacing. Multiple item types are available, including technology-enhanced items, at different cognitive levels, allowing schools to create a variety of assessments to best meet their needs.

Naiku Online Test Administration and Reports

With Naiku, the benchmark assessments are delivered online in a user-friendly web application for students (see Figure 1). Note the optional features shown where students are asked to rate their confidence in their answer selection and the "Tell me more" feature which allows students to tell their teachers more about their thinking process as they solved that question. These features engage students in the test taking process; they help students turn a test-taking event into a learning event.
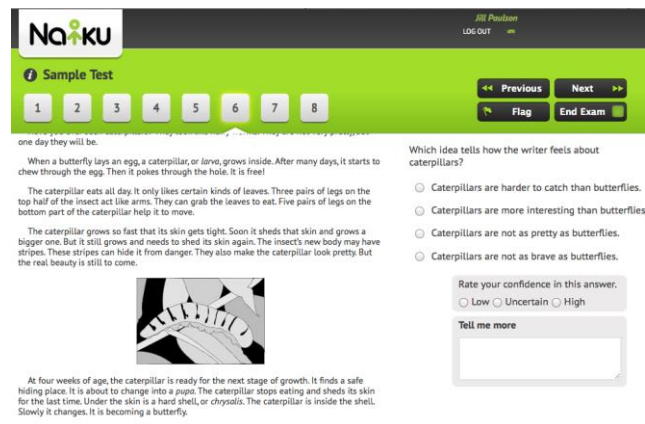

Figure 1. Naiku Test-Taking Interface. A user-friendly and intuitive test-taking interface for students, allowing them multiple ways to navigate through the test.

Naiku includes a full scoring and reporting suite. Naiku provides live, built-in reports that give clear and actionable information to inform instructional decisions. Upon test completion, individual test and longer-term longitudinal reports are immediately available to students, teachers, and staff. These reports are customizable in real-time so teachers can zero in on specific performance or specific students. The reports do not need to be "run", but the platform immediately responds to teacher input and updates the report on-the-fly.

Reports in Naiku provide educators with information to help evaluate the quality of the assessments and items. The item analysis report includes measures of item difficulty and item discrimination. The test analysis report includes an internal consistency coefficient of reliability. Class performance reports provide results disaggregated by academic standards or learning targets. These reports provide educators with the information necessary to evaluate the validity and reliability of their assessments. These reports and statistical calculations are provided for both pre-built assessments and newly-developed assessments.

A sample test analysis report is shown in Figure 2. Note the reliability estimate in the Assessment Statistics box. This statistic is the Coefficient Alpha, which provides a good estimate of the reliability of the assessment. Also note the other descriptive statistics such as the minimum score, maximum score, mean score and standard deviation, which give a good picture of how your students performed on the assessment. At the

bottom of the report, note the scores are presented by standard (or learning target). This helps teachers identify their students' strengths and areas in need of improvement and perhaps extra instruction.

Figure 2. Sample Naiku Test Analysis Report. This report shows how the students in the class performed on the test, by standard. Students are automatically grouped into proficiency levels useful for differentiation practices.

In summary, benchmark assessments occupy an important space in a comprehensive assessment system. They provide immediate results that impact classroom instruction and educational programs and policy. It is beneficial to give periodic benchmark assessments,

including up to once a month. Naiku is equipped with pre-built benchmark assessments from Inspect. With the included item bank, schools can also develop their own benchmark assessments. Reports from Naiku help teachers and administrators make informed instructional and policy decisions.

References

Black, P. & Wiliam, D. (1998). Inside the Black Box: Raising standards through classroom assessment, School of Education, King's College, London, United Kingdom.

Data Use (n.d). Assessment Systems: Benchmark Role. Retrieved from http://datause.cse.ucla.edu/ba_comprehensive.php

Data Use (n.d). Purpose of Benchmark Assessments: Retrieved from http://datause.cse.ucla.edu/ba_purpose.php

Data Use (n.d). Selecting and Using Benchmark Assessments. Retrieved from http://datause.cse.ucla.edu/ba_criteria.php

About the Author

Dr. Adisack Nhouyvanisvong is an expert in educational assessment, including computer-based and adaptive testing. He has created and ensured the psychometric integrity of large-scale educational assessments for states and national organizations, taught at the University of Minnesota, and is an adjunct professor at St. Mary's University where he has taught educators and graduate students in educational assessment practice and instructional strategies. He has been published in peer-reviewed journals, regularly speaks at education conferences and is currently President of Naiku.