



Five Features of High-Quality Assessment Content

How to evaluate formative assessment content
to obtain more meaningful student feedback

CONTENTS

- 3 Flexibility
- 4 Validity
- 4 Reliability
- 5 Variety
- 6 Insight
- 7 The Bottom Line: Driving Improvement



Five Features of High-Quality Assessment Content

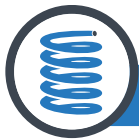
How to evaluate formative assessment content to obtain more meaningful student feedback

Teachers need reliable assessment data to guide their instruction and make sure that every child meets rigorous learning standards. But the data they get from a formative or interim assessment is only as good as the quality of the instrument itself.

Without accurate information from an assessment, educators can misdiagnose a child's learning needs and spend time focusing on the wrong concepts as a result.

Creating high-quality assessment forms (tests) and items (questions) takes time and expertise. Not all assessment content will produce valid results.

To make sure the interim and formative assessment content they are purchasing for their schools provides accurate information to support high-quality instruction, K-12 leaders should look for these five key elements.



FLEXIBILITY

Does the assessment content include options for educators to give pre-constructed benchmark tests, as well as create their own exams by choosing questions and performance tasks from among an item bank? Does it allow educators to give either computer-adaptive or fixed-format assessments? What about paper-based or computer-based formats?

There are advantages and limitations to all these various approaches. High-quality assessment content should give educators a choice in which approach they would like to take for different exams, based on their needs and goals.



Computer adaptive tests are useful for quickly homing in on a student's precise skill level or determining gaps in skills.

According to Mel Lee, product manager for assessment, benchmark tests, such as the K-12 Grade Level Readiness exams from McGraw-Hill Education, allow school districts to test how ready students are for the content they are about to receive in the grade level they are about to enter. Educators can determine how likely students are to achieve grade-level standards, and they can even compare a student's achievement to national norms with a percentile score. The tradeoff is that educators don't have control over the individual items in these assessments.

To give educators this degree of control, assessment platforms should allow users to build their own interim or formative assessments using a comprehensive item bank, and this is exactly what assessment technology from McGraw-Hill Education does. This way, educators can build customized assessments that map directly to their curriculum.

In a fixed-form test, every student receives the same items. With an adaptive test, a computer algorithm scores each item or set of items as the student answers, then adjusts the level of the next items depending on the response. Because the level of difficulty adjusts on the fly to the level demonstrated by the student, students are consistently challenged without being overwhelmed by items that are too hard or bored by items that are too easy.

Computer adaptive tests are useful for quickly homing in on a student's precise skill level or determining gaps in skills. However, because each student might receive different test items, adaptive tests cannot be used for comparing the achievement of students to a national average or to a group of their peers.

Having a variety of options in the kinds of assessments they can give allows educators to choose the best format to meet their specific goals for each exam.



VALIDITY

Does the assessment content provide a valid way to measure the skills you are looking to assess? Simply put, does it test what you want to test?

Not only must the content be tied to high-quality learning standards, but each item must accurately measure the skill or standard to which it correlates. To ensure the validity of test items, assessment content should be constructed using a rigorous, evidence-based design process that asks: How can students demonstrate this particular skill? And, how can a test item prove it?

Thousands of hours of research and field testing go into the benchmark tests and item banks offered by McGraw-Hill Education. “We have content experts look at our assessment items and make sure they are measuring what they should measure,” says Lee. Only items that pass this rigorous evaluation process make it into the item banks for use.



RELIABILITY

Assessment forms and items also should be tested by researchers to ensure their reliability. In other words, if a student took the same exam over and over again, would the results be consistent across these multiple attempts?

All McGraw-Hill Education content is rigorously evaluated for its reliability as well as its validity. “We look at factors such as *P* value, or the proportion of students who answer an item correctly,” says Lee. “We send any items that have statistical anomalies to content experts, and ask why they think those items aren’t performing consistently. We also look at how high-achieving students perform on items compared with low-achieving students. If low-achieving students perform well but high-achieving students don’t, then we know there must be something wrong with that item.”

The Penn-Harris-Madison School Corporation in Indiana uses assessments from McGraw-Hill Education for giving benchmark exams and for creating other interim assessments to inform instruction. Assistant Superintendent for Instruction Kay Antonelli says the content from McGraw-Hill Education is extremely reliable as a predictor of student success.

“We did a correlation between how our students in grades K–8 scored on the Form C benchmark exams and our state standardized assessment,” she says. “We wanted to see whether Form C was a good predictor of how our students would fare on the summative state assessment. We found that for mathematics, the correlation was 0.86, which is extremely high, and for ELA it was 0.78. The process showed that the K–12 Readiness Assessments are a strong indicator of our students’ capability in mastering essential standards.”



VARIETY

Does the assessment content provide a wide variety of item types, from selected response questions to technology-enhanced items and performance tasks? Can educators give exams with questions that span a broad range of complexity and difficulty levels?

Different kinds of assessment questions serve different purposes. “Selected response items often get a bad rap,” says Lee, “but they can serve a valid purpose.” If you are looking to test students’ recall or their grasp of basic skills and concepts, then selected response questions might be appropriate—and they are also easy to score. However, as you move further up the Depth of Knowledge (DOK) scale, which is something that next-generation learning standards require, selected response items are no longer sufficient.

“If you want to see evidence that students can take multiple source documents, compare and contrast them, and form an argument, then a writing prompt is a more valid way to assess those skills,” he says.

The key to eliciting meaningful information about students’ skills and abilities is to match the kinds of questions you are asking with the DOK you are looking to measure—and having a wide variety of item types to choose from allows you to do this effectively. “Test authors can pick the right item type to match the task or construct they want to measure,” notes Lee.



The key to eliciting meaningful information about students’ skills and abilities is to match the kinds of questions you are asking with the DOK you are looking to measure—and having a wide variety of item types to choose from allows you to do this effectively.

According to Lee, to evaluate students’ learning at a deeper level, students must produce something—such as an essay or a constructed response to a math problem. But those kinds of items can be expensive to grade. Technology-enhanced items let educators test students at greater depths of knowledge while still allowing for the items to be auto-scored, which saves teachers time and effort.

Penn-Harris-Madison trains teachers how to align the kinds of questions they ask on assessments to the DOK they want students to demonstrate. “Having a variety of question types is essential,” says Antonelli. “And without technology-enhanced items, I think we would be relegated to testing DOK 1 or 2, where we’re asking students to give us an answer. That’s not really interpretive of higher levels of thinking.”



INSIGHT

Finally, does the assessment content include reports that provide actionable data for teachers and administrators? The whole purpose of formative assessment is to help educators become more effective by using the results to target and improve their instruction—and reports that provide key insights make this process easier.

Assessments from MHE include reports to provide valuable insight to all stakeholders, from educators and administrators to parents and students. These reports, and the detailed information behind them, can be used to inform classroom instruction, modify teaching strategies, and assess progress. Administrators can monitor growth and performance at the student, grade, school, district, and state level and evaluate overall program effectiveness. Teachers can gain insight at the student level to improve lessons and collaborate on effective strategies. Students and parents can use reports to track specific learning goals and overall progress as well, which helps put students in charge of their own learning.

“After every formative assessment, we look closely at the data, and the reports are really informative for us,” says Antonelli. “We can look at an individual student clear down to the areas where he or she is struggling. But the types of reports that MHE provides are multidimensional, so we can also look across each of our classes, across schools, and across the district. This allows us to see if there are patterns.”

She adds: “Say, for example, all of the children in fifth grade are having difficulty with one particular math standard. We might conclude that there is a piece of our curriculum map that we need to change, because we’re not covering that standard quickly enough or deeply enough for students to be able to show they understand it. We use the information in these reports in a very broad way across the district, so that we can make constructive changes.”



The Bottom Line: Driving Improvement

High-quality assessment content (1) allows educators to match the type of assessment they give with the goals they hope to accomplish, and (2) delivers accurate information that leads to deeper insight into students' strengths, weaknesses, and areas for improvement.

To meet both of these needs, K-12 leaders should look for assessment content that can be delivered flexibly, with a wide variety of question types—as well as content that is engineered to produce valid and reliable results. In addition, the content should feature in-depth reporting to give educators a clear understanding of students' growth and abilities.

“Our use of formative assessment gives us information to help student’s master grade-level content, and it helps our teachers provide more targeted instruction in their classrooms,” says Antonelli. “We know that formative assessment is absolutely essential in the learning cycle, and we rely on assessment content and technology from McGraw-Hill Education to help us collect reliable, actionable information that drives improvement.”



“Our use of formative assessment gives us information to help students master grade-level content, and it helps our teachers provide more targeted instruction in their classrooms.”



About McGraw-Hill Education

Our vision is to unlock the full potential of each learner.

Our mission is to accelerate learning through intuitive, engaging, efficient, and effective experiences—grounded in research.

At McGraw-Hill Education, we believe that our contribution to unlocking a brighter future lies within the application of our deep understanding of how learning happens and how the mind develops. It exists where the science of learning meets the art of teaching.

Educators have been and always will be at the core of the learning experience. The solutions we develop help educators impart their knowledge to students more efficiently. We believe that harnessing technology can enhance learning inside and outside of the classroom and deepen the connections between students and teachers to empower greater success.

By partnering with educators around the globe, our learning engineers, content developers, and pedagogical experts are developing increasingly open learning ecosystems that are proven to improve pass rates, elevate grades, and increase engagement for each individual learner while improving outcomes for all.

Why? Because learning changes everything.

Visit www.mheducation.com

eSCHOOL NEWS

This white paper was produced by eSchool News, a leading publisher and producer of K-12 educational technology publications and events, dedicated to the advancement and wise use of technology to improve teaching and learning for all. eSchool News offers ed-tech decision-makers a range of products—including magazines, white papers, websites, newsletters, webinars, and other products—that provide in-depth coverage of the latest innovations, trends, and real-world solutions impacting the education community. Explore more at <http://www.eSchoolNews.com>.