

Instructionally-Sensitive Assessments

: 11/28/2018

Across the country educators are evaluating current instructional practices and implementing strategies designed to improve student learning. To evaluate students' progress, educators administer multiple tests each year, including district-level interim and benchmark assessments and state-level annual summative assessments. Educators are using test results, along with other information gathered during instruction, to gauge whether changes made to instructional design and practice are working.

James Popham (2007) identified a key premise in using test results to inform instructional practice. "The premise underlying the use of these accountability tests is that students' test scores will indicate the quality of instruction those students have received." However, in most cases, there is little evidence that test scores should be used to evaluate the quality of instruction that students received (Naumann, Hochweber, & Klieme, 2016). To address the connection between a student's performance on a test and the quality of instruction, researchers are developing methods to evaluate the instructional sensitivity of a test or item. According to Popham (2007), "[a] test's *instructional sensitivity* represents the degree to which students' performances on that test accurately reflect the quality of the instruction that was provided specifically to promote students' mastery of whatever is being assessed."

Researchers are exploring two primary approaches to evaluating the extent to which a test or test item is instructionally sensitive: expert judgment and psychometric analysis.

Expert Judgment

The first approach using expert judgment, detailed by Popham (2007), asks educators to use rubrics to evaluate four questions related to instructional sensitivity:

1. To what degree can teachers deliver instruction aligned to the curricular aims in the time allowed?
2. To what degree do teachers understand the knowledge and skills to be assessed?
3. Are there enough items to justify the claims being made, and to what degree is the domain being evaluated?
4. To what degree are the items on the test judged to be sensitive to instructional impact?

Psychometric Analysis

With the psychometric approach, researchers begin by empirically determining the amount of variance in an item or test form and then connect the variance to empirical measures of instructional quality. Nauman, Hochweber, and Klieme (2016) describe three models for determining the amount of variance in an item or

test: a differences-between-groups model, a differences-between-time-periods model, and a model that combines the two. Once the variances have been calculated, the results are correlated to empirical measures of instructional quality. The results of these analyses show the extent to which instructional quality is the likely culprit of variance and shows the extent to which other culprits of variance can be ruled out.

Empirical Measures of Instructional Sensitivity

Researchers are using empirical measures of instructional sensitivity that rely on survey and observational data. Two instructional-sensitivity evaluations, one conducted by D’Agostino, Welsh, and Corson (2007) and one conducted by Polikoff (2016), both used survey data to develop an instructional-quality measure that could be correlated to variances in student performance on the tests.

D’Agostino, Welsh, and Corson (2007) developed a survey, administered to teachers, containing a series of open-ended questions that collected information about classroom instruction. Subject matter experts used rubrics to review each survey response. From the data, the researchers developed an alignment index, which was used in their instructional-sensitivity evaluation as the index of instructional quality. Higher scores on the alignment index indicated “more commonality between how the test and teacher operationally defined the [performance objectives].” Their results showed that, “[teachers] who reported greater standards emphases and whose teaching matched the test had greater adjusted scores, on average.” Students benefitted from learning the standards “similar to the way they were being tested.”

Polikoff (2016) used survey and observational data from the Bill and Melinda Gates Foundation’s Measures of Effective Teaching (MET). The study included data from “multiple survey and observational measures at the class-section level in each of two years”. The data were used to develop multiple value-added measures that were correlated to student performance on the state’s summative test. Polikoff conducted these analyses for four state summative tests, and the results showed that “most of the state assessments showed a modest sensitivity to one or more of the observational measures of effective teaching.”

How Can This Information be Used?

Evaluating the instructional sensitivity of items and tests provides a couple of important benefits. First, the data from instructionally-sensitive tests can be incorporated into a school or district’s evaluation of instructional practice. Educators will be able to assess the degree to which the results can be used for this purpose, leading to decisions that will improve student learning. Second, educators will be more motivated to support testing, seeing tests as connected to their work rather than as an activity separate from instruction. When educators see the value in using the results to inform instructional practice, they will be propelled to use the data to make informed choices. Educators and policymakers can maximize the effectiveness of large-scale testing to inform instructional design and practice, and ultimately improve student learning, by evaluating the instructional sensitivity of tests and test items.

References

D'agostino, J. V., Welsh, M. E., & Corson, N. M. (2007). Instructional sensitivity of a state's standards-based assessment. *Educational Assessment*, 12(1), 1–22. <https://doi.org/10.1080/10627190709336945>

Naumann, A., Hochweber, J., & Klieme, E. (2016). A psychometric framework for the evaluation of instructional sensitivity. *Educational Assessment*, 21(2), 89–101. <https://doi.org/10.1080/10627197.2016.1167591>

Polikoff, M. S. (2010). Instructional sensitivity as a psychometric property of assessments: winter, 2010. *Educational Measurement: Issues and Practice*, 29(4), 3–14. <https://doi.org/10.1111/j.1745-3992.2010.00189.x>

Polikoff, M. S. (2016). Evaluating the instructional sensitivity of four states' student achievement tests. *Educational Assessment*, 21(2), 102–119. <https://doi.org/10.1080/10627197.2016.1166342>

Popham, W. J. (2007). Instructional insensitivity of tests: accountability's dire drawback. *Phi Delta Kappan*, 89(2), 146–155. <https://doi.org/10.1177/003172170708900211>

Chad M. Barrett is senior advisor at AdvancED | Measured Progress. In this role Mr. Barrett crafts solutions strategies and proposal text as part of the company's statewide large-scale assessment work. He has also worked with state education departments to manage assessment programs and improve the technical quality of these programs. Mr. Barrett started in educational assessment as a content development specialist. In this role he edited test items, verified the alignment of items to content and cognitive complexity standards, and facilitated item review and bias review committees. Prior to his career in educational assessment, he taught middle school and high school mathematics and served as the principal for a private school that worked with at-risk youth. Mr. Barrett earned a master degree in classroom education from Walden University. He earned his bachelor degree from Ithaca College, where he completed the teacher preparation program in middle and high school mathematics.

© Cognia Inc.

This article may be republished or reproduced in accordance with [The Source Copyright Policy](#).

