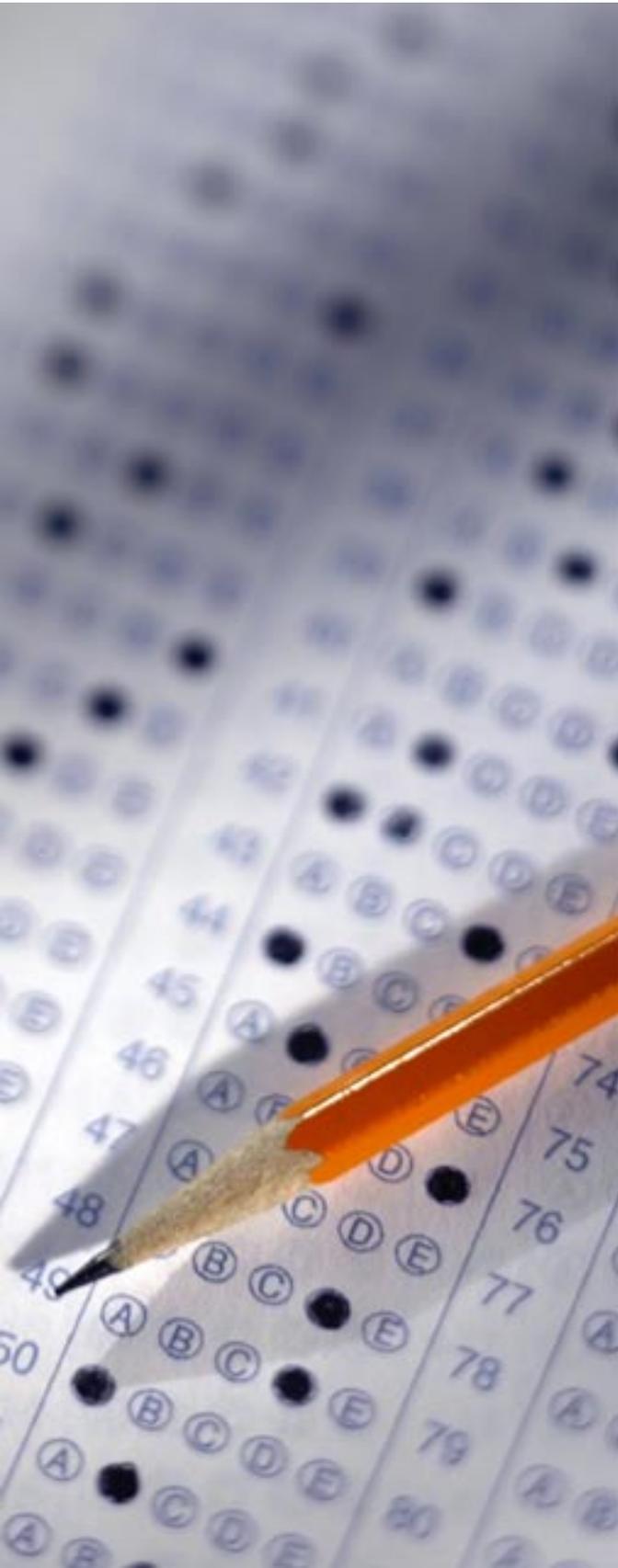




The predictive validity of selected benchmark assessments used in the Mid-Atlantic Region



Institute of Education Sciences
U.S. Department of Education



The predictive validity of selected benchmark assessments used in the Mid-Atlantic Region

Summary

November 2007

Prepared by

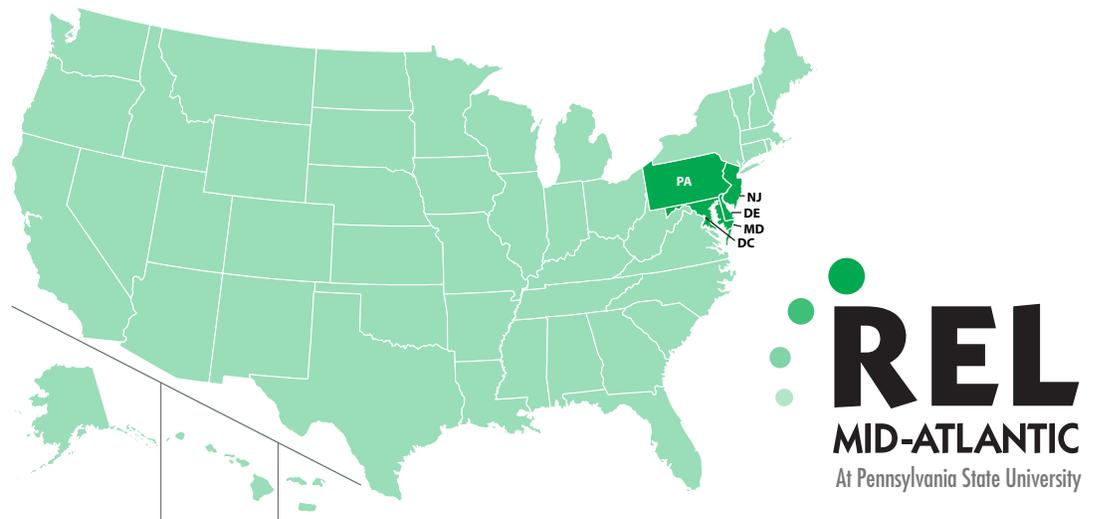
**Richard S. Brown
University of Southern California**

**Ed Coughlin
Metiri Group**



Institute of Education Sciences

U.S. Department of Education



Issues & Answers is an ongoing series of reports from short-term Fast Response Projects conducted by the regional educational laboratories on current education issues of importance at local, state, and regional levels. Fast Response Project topics change to reflect new issues, as identified through lab outreach and requests for assistance from policymakers and educators at state and local levels and from communities, businesses, parents, families, and youth. All Issues & Answers reports meet Institute of Education Sciences standards for scientifically valid research.

November 2007

This report was prepared for the Institute of Education Sciences (IES) under Contract ED-06-CO-0029 by Regional Educational Laboratory Mid-Atlantic administered by Pennsylvania State University. The content of the publication does not necessarily reflect the views or policies of IES or the U.S. Department of Education nor does mention of trade names, commercial products, or organizations imply endorsement by the U.S. Government.

This report is in the public domain. While permission to reprint this publication is not necessary, it should be cited as:

Brown, R. S., & E. Coughlin. (2007). *The predictive validity of selected benchmark assessments used in the Mid-Atlantic Region* (Issues & Answers Report, REL 2007–No. 017). Washington, DC: U.S. Department of Education, Institute of Education Sciences, National Center for Education Evaluation and Regional Assistance, Regional Educational Laboratory Mid-Atlantic. Retrieved from <http://ies.ed.gov/ncee/edlabs>

This report is available on the regional educational laboratory web site at <http://ies.ed.gov/ncee/edlabs>.

Summary

The predictive validity of selected benchmark assessments used in the Mid-Atlantic Region

This report examines the availability and quality of predictive validity data for a selection of benchmark assessments identified by state and district personnel as in use within Mid-Atlantic Region jurisdictions. The report finds that evidence is generally lacking of their predictive validity with respect to state assessment tests.

Many districts and schools across the United States have begun to administer periodic assessments to complement end-of-year state testing and provide additional information for a variety of purposes. These assessments are used to provide information to guide instruction (formative assessment), monitor student learning, evaluate teachers, predict scores on future state tests, and identify students who are likely to score below proficient on state tests.

Some of these assessments are locally developed, but many are provided by commercial test developers. Locally developed assessments are not usually adequately validated for any of these purposes, but commercially available testing products should provide evidence of validity for the explicit purposes for which the assessment has been developed (American Educational Research Association, American Psychological Association, & National Council

on Measurement in Education, 1999). But the availability of such information and its interpretability by district personnel vary across instruments. When the information is not readily available, it is important for the user to establish such evidence of validity. A major constraint on district testing programs is the lack of resources and expertise to conduct validation studies of this type.

As an initial step in collecting evidence on the validity of district tests, this study focuses on the use of benchmark assessments to predict performance on state tests (predictive validity). Based on a review of practices within the school districts in the Mid-Atlantic Region, this report details the benchmark assessments being used, in which states and grade levels, and the technical evidence available to support the use of these assessments for predictive purposes. The report also summarizes the findings of conversations with test publishing company personnel and of technical reports, administrative manuals, and similar materials.

The key question this study addresses is: What evidence is there, for a selection of commonly used commercial benchmark assessments, of the predictive relationship of each instrument with respect to the state assessment?

The study investigates the evidence provided to establish a relationship between district and state test scores, and between performance on district-administered benchmark assessments and proficiency levels on state assessments (for example, at what cutpoints on benchmark assessments do students tend to qualify as proficient or advanced on state tests?). When particular district benchmark assessments cover only a subset of state test content, the study sought evidence of whether district tests correlate not only with overall performance on the state test but also with relevant subsections of the state test.

While the commonly used benchmark assessments in the Mid-Atlantic Region jurisdictions may possess strong internal psychometric characteristics, the report finds that evidence is generally lacking of their predictive validity with respect to the required state or summative assessments. A review of the evidence for the four benchmark assessments considered—Northwest Evaluation Association’s Measures of Academic Progress (MAP; Northwest Evaluation Association, 2003), Renaissance Learning’s STAR Math/STAR Reading (Renaissance Learning, 2001a, 2002), Study Island’s Study Island (Study Island, 2006a), and CTB/McGraw-Hill’s TerraNova (CTB/McGraw-Hill, 2001b)—finds documentation of criterion validity of some sort for three of them (STAR, MAP, and TerraNova), but only one was truly a predictive study and demonstrated strong evidence of predictive validity (TerraNova).

Moreover, nearly all of the criterion validity studies showing a link between these benchmark assessments and state test scores in the Mid-Atlantic Region used the Pennsylvania State System of Assessment (CTB/McGraw-Hill, 2002a; Renaissance Learning, 2001a, 2002) as the object of prediction. One study used the Delaware Student Testing Program test as the criterion measure at a single grade level, and several studies for MAP and STAR were related to the Stanford Achievement Test–Version 9 (SAT–9) (Northwest Evaluation Association, 2003, 2004; Renaissance Learning, 2001a, 2002) used in the District of Columbia. None of the studies showed predictive or concurrent validity evidence for tests used in the other Mid-Atlantic Region jurisdictions. Thus, no predictive or concurrent validity evidence was found for any of the benchmark assessments reviewed here for state assessments in Maryland and New Jersey.

To provide the Mid-Atlantic Region jurisdictions with additional information on the predictive validity of the benchmark assessments currently used, further research is needed linking these benchmark assessments and the state tests currently in use. Additional research could help to develop the type of predictive validity evidence school districts need to make informed decisions about which benchmark assessments correspond to state assessment outcomes, so that instructional decisions meant to improve student learning as measured by state tests have a reasonable chance of success.

November 2007