

Center for Educational Assessment (CEA) MCAS Validity Studies
Prepared By
Center for Educational Assessment
University of Massachusetts Amherst

All of the following CEA MCAS Validity Reports are available at the website:
<http://www.umass.edu/remap>. Many of the reports include abstracts. For each report, a brief statement about the contents has been provided.

Here is an example (using the first report) for how the reports can be cited:

Zenisky, A. L., & Hambleton, R. K. (2003). **DIF Detection and Interpretation in Large-Scale Science Assessments: Informing Item-Writing Practices** (Center for Educational Assessment MCAS Validity Report No. 1). Amherst, MA: University of Massachusetts, Center for Educational Assessment.

CEA MCAS Validity Report No. 1. (CEA-429). DIF Detection and Interpretation in Large-Scale Science Assessments: Informing Item-Writing Practices. April L. Zenisky, Ronald K. Hambleton. March 17, 2003.

This study investigated Differential Item Functioning (DIF) in the MCAS Science and Technology assessments over two years and three grade levels. Several findings that could be useful to test development committees emerged from the study. For example, when there is DIF in science and technology items, multiple-choice items tend to favor males and open-response items tend to favor females.

CEA MCAS Validity Report No. 2. (CEA-454). Evaluating the Consistency of Test Content across Two Successive Administrations of a State-Mandated Science and Technology Assessment. Tim O’Neil, Stephen G. Sireci, Kristen Huff. April 2002.

Tests used for accountability such as the MCAS tests must demonstrate content validity and be consistent from year to year. In this study, the researchers investigated the grade 10 Science and Technology assessments over two years. Their findings are quite positive regarding content validity and have implications for the future design of assessments.

CEA MCAS Validity Report No. 3. (CEA-458). Comparing the Psychometric Properties of Monolingual and Dual Language Test Forms. Stephen G. Sireci, S. Khaliq. June 2002.

Many students in the United States who are required to take educational tests are not fully proficient in English. To address this problem, a state-mandated testing program in Massachusetts created dual language English-Spanish versions of tests other than the ELA tests. In this study, the structural equivalence of the Spanish and English versions of the MCAS grade 4 Mathematics test was analyzed, and a DIF study was conducted. The findings show a high comparability of the two language versions and a minimal level of DIF.

CEA MCAS Validity Report No. 4. (CEA-477). Student Test Score Reports and Interpretive Guides: Review of Current Practices and Suggestions for Future Research. Dean P. Goodman, Ronald K. Hambleton. June 19, 2003.

In this study, the researchers investigated the student reports and interpretative guides from 11 states, 2 Canadian provinces, and 3 major national test publishers. Based on their review, they offered a set of recommendations to practitioners for improving student reports. These recommendations are intended to help state departments of education avoid some of the many pitfalls of test score report design.

CEA MCAS Validity Report No. 5. (CEA-531). A Multitrait-Multimethod Validity Investigation of the 2002 Massachusetts Comprehensive Assessment System Tests. Dean Goodman. April 30, 2004.

In this study, the researcher compiled construct validity evidence (using the multitrait-multimethod approach) on the grades 4 and 10 ELA and Mathematics assessments. The results are supportive of the construct validity of these four tests and highlight the conclusion that in both subject areas and at both grade levels the multiple-choice and open-response items were very highly correlated and measured the relevant constructs.

CEA MCAS Validity Report No. 6. (CEA-537). Comparison of Trends in NAEP, Massachusetts-NAEP, and MCAS Results. Stephen Jirka, Ronald K. Hambleton. July 4, 2004.

The purpose of this study was to compare average proficiency levels, proficiency classifications, and trends between National Assessment of Educational Progress (NAEP) results for the country and NAEP results for Massachusetts, and then NAEP results for Massachusetts and MCAS results at grades 4 and 8. In general, Massachusetts outperformed the nation in both reading and mathematics. The MCAS mathematics results were very much in line with the trends seen in the NAEP results for Massachusetts. On the other hand, MCAS ELA results were a bit out of line with the NAEP state results, with the MCAS ELA results showing larger gains. Further investigation of the NAEP and MCAS proficiency descriptions and achievement levels could be revealing and help in the interpretation of the ELA results. Cautions in making the comparisons are highlighted in the report.

CEA MCAS Validity Report No. 7. (CEA-538). Alignment of MCAS Grade 10 English Language Arts and Mathematics Assessments with the Massachusetts Curriculum Frameworks and the Test Specifications. Ronald K. Hambleton, Yue Zhao. December 18, 2004.

Alignment studies among test content, test content specifications, and state curriculum frameworks became extremely important as states gathered evidence for the USDOE peer review process to meet the requirements of No Child Left Behind (NCLB). This was the first of several studies carried out in Massachusetts to address questions regarding alignment. The report clearly demonstrates that the MCAS grade

10 ELA and Mathematics assessments from 1998–2004 showed diversity of content and content that closely matched the test specifications and curriculum frameworks. In both ELA and mathematics, the test content specifications were met almost perfectly (especially in the last four years), and all, or nearly all, of the learning standards have been included in the assessments.

CEA MCAS Validity Report No. 8. (CEA-540). MCAS 2001 Grade 10 ELA and Mathematics Model Fit Analyses. Ning Han. December 8, 2003.

The purpose of this study was to apply a new method to model fit by checking the predicted score distribution and the observed one to assess the fitness of Item Response Theory (IRT) models to MCAS data. The study was carried out with the grade 10 assessments because these two assessments are the most consequential to students. The report of the study’s findings shows that the grade 10 ELA and Mathematics assessments were unidimensional, and the predictions of test score distributions were excellent with several IRT models, including the models then in use by the state.

CEA MCAS Validity Report No. 9. (CEA-558). Evaluating the Fit between Test Content, Instruction, and Curriculum Frameworks: A Review of Methods for Evaluating Test Alignment. Shuhong Li, Stephen G. Sireci. June 30, 2005.

Content validity has long been recognized as the cornerstone of a sound educational assessment, and the effort invested in achieving and evaluating content validity has been ongoing. In this report, the existing literature on test alignment was reviewed, and the five most popular methods were evaluated using several criteria. The implications of the study are relevant to test developers and local and state education agencies.

CEA MCAS Validity Report No. 10. (CEA-566). Cognitive Complexity Levels for the MCAS Assessments. Stephen Jirka, Ronald K. Hambleton. June 1, 2005.

This brief report provides some background information regarding several of the various methods for assessing cognitive complexity that have been devised and offers a recommendation for a model that would be appropriate to use on items found in the MCAS tests.

CEA MCAS Validity Report No. 11. (CEA-575). Test-Curriculum Alignment Study for MCAS Grades 4 and 7 ELA and Grades 4, 6, and 8 Mathematics. Ronald K. Hambleton, Yue Zhao. November 24, 2005.

Two of the technical requirements for valid state testing programs are (1) the content of the tests must be consistent with expectations—that is, the content specifications prepared for the tests, and (2) the tests must show content diversity over time. This study assessed the extent to which the two requirements above, which are part of the technical requirement known as “content validity,” were met by the grades 4 and 7

ELA, and grades 4, 6, and 8 Mathematics assessments administered between 2001 and 2004. The research findings suggest that the matches between the ELA and Mathematics test content specifications and the actual assessments were close for nearly all of the assessments constructed between 2001 and 2004.

CEA MCAS Validity Report No. 12. (CEA-576). Alignment of the MCAS Assessments to State Content Standards: An Example. Stephen J. Jirka, Ronald K. Hambleton. September 30, 2005.

This brief report provides examples of the types of evidence states might gather to show that they are compliant with the federal legislation regarding alignment. The report provides some background information regarding (1) why it is necessary to have an independent alignment study, (2) what method might be used, (3) a sample protocol for an alignment study, and (4) what type of data such a study might provide. The goal of this report was to offer a methodology that extends the work of Hambleton and Zhao (see CEA MCAS Validity Report No. 11) to the grade 10 ELA and Mathematics assessments and is consistent with federal guidelines.

CEA MCAS Validity Report No. 13. (CEA-584). Vertical Scaling for the MCAS. Craig Wells, Lisa Keller. August 30, 2005.

This report examines several issues related to developing a vertical scale for the MCAS ELA and Mathematics assessments. First, a general vertical scaling design was proposed. Second, the feasibility of establishing a vertical scale for the MCAS ELA and Mathematics assessments was addressed in light of the proposed design.

CEA MCAS Validity Report No. 14. (CEA-585). Investigation of MCAS-NAEP Comparisons and Other External Validity Evidence. Stephen J. Jirka, Ronald K. Hambleton. June 30, 2005.

This report provides evidence compiled by the researchers about the validity of the MCAS tests in light of data that are external to the assessments themselves: from the National Assessment of Educational Progress (NAEP), standardized achievement tests, Achieve Inc. data, and Keep the Promise data. For each area of evidence, the researchers provided the following information: background and goals of the study or studies, methodology, main results, and conclusions. The findings suggest strong evidence for MCAS validity.

CEA MCAS Validity Report No. 15. (CEA-612). MCAS 2006 Curriculum-Test Alignment Study Update. Ronald K. Hambleton, Yue Zhao, Zachary R. Smith. November 10, 2006.

In 2004 and 2005 Hambleton and Zhao carried out curriculum-test alignment studies for grades 4, 7, and 10 English Language Arts and grades 4, 6, 8, and 10 Mathematics for the time period between 2001 and 2004, and at the grade 10 level from 1998 to 2004. (See CEA MCAS Validity Reports No. 7 and 11.) Two technical criteria were used in the evaluation of the curriculum-test alignments: (1) the degree of match between the content specifications for the tests and the actual percentage of score points included for each content strand, and (2) the percentage of learning standards

included in the tests over each three-year period since 2001. The purpose of this report was to bring the curriculum-test alignment evidence for the seven tests up-to-date by reporting on the findings from 2005 and 2006 in the context of the curriculum-test alignments over the last six years. The results are similar to those presented in earlier studies and show a very high curriculum alignment for all the updated tests.

CEA MCAS Validity Report No. 16. (CEA-613). Test-Curriculum Alignment Study for the Grade 3 Reading Test and the New MCAS tests in 2006: Grades 5, 6, and 8 English Language Arts, and Grades 3, 5, and 7 Mathematics. Ronald K. Hambleton, Zachary R. Smith, Yue Zhao. November 10, 2006.

The intent of the contractor and the Department of Education is to build MCAS tests each year that (1) are in alignment with the test content specifications, (2) over regular intervals of time assess all of the learning standards in each curriculum framework that are intended to be included in the tests, and (3) use test items that are valid indicators of the learning standards to which they are matched. This report presents the findings from studies that assessed the extent to which the first requirement above, which is one of the technical requirements for curriculum-test alignment, is met by the grade 3 Reading; grades 5, 6, and 8 English Language Arts; and grades 3, 5, and 7 Mathematics tests administered in 2006. The research findings are clear: For all six new tests and the grade 3 Reading test, for which a new performance level was set in 2006, the actual distribution of test content was in nearly perfect alignment with the test content specifications. This is an excellent result. The researchers also determined that the contractor and the Department took an excellent first step toward meeting requirement two by assessing relatively large percentages of the learning standards in the first year of testing, and certainly as large as the research team observed in other subject areas and grade levels with a longer history than that of the six new tests studied in this report. Only one recommendation seemed necessary. The Department and contractor need to continue to monitor the degree to which the learning standards intended for classroom assessment are actually being assessed at the classroom level.

CEA MCAS Validity Report No. 17. (CEA-649). Psychometric Analyses of the 2006 MCAS High School Science and Technology/Engineering Tests. Ronald Hambleton, Yue Zhao, Zachary Smith, Wendy Lam, Nina Deng. January 30, 2008.

The primary goals of this study are to (1) determine the psychometric similarities and differences among the four new high school Science and Technology/Engineering (STE) tests, and (2) provide worthwhile psychometric data on each test that might help in the evaluation and ongoing development of these tests. These goals are consistent with the NCLB legislation that requires states to “have implemented a set of high quality, yearly student assessments...in science” (NCLB, 2001) with the focus on the psychometric quality of the tests.

Extensive research showed the four STE tests in their current form are sound technically, and highly comparable in quality. The presence of a very small number

of DIF items and some less than optimal placements of test information functions were small flaws in the overall, excellent quality of the tests. Four supplements to this report are also available in printed copies and provide substantial details on the analyses and specific findings for each of the four subjects. (See page 7 of this listing for titles and dates.)

CEA MCAS Validity Report No. 18. (CEA-689). Redesign of MCAS Tests Based on a Consideration of Information Functions. (Revised Version). Ronald K. Hambleton, Wendy Lam. January 9, 2009

In this study, the researchers investigated the impact of raising item discrimination indices on the validity of performance category assignments. The report of their findings shows that the current ELA and mathematics Test Information Functions (TIFs) in grades 6 and 10, met or exceeded TIF expectations at the three cut scores (with one exception, grade 6 ELA at the *Advanced* performance level.). The report also shows the benefits of increasing the average discriminating power of the test items. How practical this might be remains to be seen. Certainly a review of current item writing practices and review procedures and a review of the best test items (to spot features that might be repeated in more items) could be initiated to explore possible options.

CEA MCAS Validity Report No. 19. (CEA-690). MCAS Equating Research Report: An Investigation of FCIP-1, FCIP-2, and Stocking and Lord Equating Methods. Lisa A. Keller, Ronald K. Hambleton, Pauline Parker, Jenna Copella. December 10, 2008.

Three studies were conducted to compare three equating methods: two implementations of fixed common item parameter equating, FCIP-1 and FCIP-2, and the Stocking and Lord test characteristic curve method. Taken together, the overall results of these three studies indicate that using either FCIP-2 or Stocking and Lord equating methods appears to lead to somewhat more accuracy than the FCIP-1 method. Since the simulations, which used grade 6 mathematics items, very closely matched the tests and the conditions in the MCAS program from year to year (with the exception of Case B in Study 3), generalization of the main findings to other grade 3 to 8 ELA and Mathematics tests seems warranted. The results from the research strongly support a recommendation to switch from the FCIP-1 to the FCIP-2 equating method.

CEA MCAS Validity Report No. 20. (CEA-709). Judging the Content and Statistical Equivalence of MCAS Operational and Linking Items. Nina Deng, Tia Sukin, Ronald K. Hambleton. May 25, 2009.

Content match between the anchor items and the operational items seems especially important because it is in the linking process that growth from one year to the next is determined. If content match (i.e., content representativeness) is not present, it would be possible to obtain a biased estimate of growth and undermine the equating process. Of the two factors, content match is more important because non-match can lead to

bias in estimates of growth. Statistical non-match can be handled fairly easily in the equating process. The researchers looked at the content match between items used in the linking process and in the operational tests and found a very high match. These findings clearly support the current approach to selecting linking items with suitable content.

The seven reports below are available in printed copy only.

These seven studies provide purposes, methodology, and results for psychometric studies that were conducted on the early editions of the MCAS—1998 and 1999. The last four of the seven highlight efforts to carry out differential item functioning (DIF) studies of gender and ethnicity (black, Hispanic, and white). The studies identified only a very small number of DIF items.

CEA-356 Psychometric Analyses of the 1998 Grade 4 Massachusetts Comprehensive Assessment System Tests. Fred Robin, Saba Rizavi, Lisa Keller, Tera Smith.

CEA-357 Psychometric Analyses of the 1998 Grade 8 Massachusetts Comprehensive Assessment System Tests. Karla Egan, Kevin Meara, Georgette Rodriguez.

CEA-358 Psychometric Analyses of the 1998 Grade 10 Massachusetts Comprehensive Assessment System Tests. B. Bastari, Dehui Xing, Alan Dillingham, Urip Purwono, April Zenisky.

CEA-364 Assessing the Dimensionality of the Grade 4 MCAS Science and Technology Test: A multimethod Analysis. Lisa Keller, Georgette Rodriguez, April Zenisky, Stephen G. Sireci.

CEA-412 Differential Item Functioning Analyses for the 2000 MCAS Grade 4 Tests. L. Ying.

CEA-413 Differential Item Functioning Analyses for the 2000 MCAS Grade 8 Tests. Sarah Klauck.

CEA-414 Differential Item Functioning Analyses for the 2000 MCAS Grade 10 Tests. X. Ma.

Supplements to CEA MCAS Validity Report No. 17. (CEA-645, 646, 647, 648)

CEA-645 Psychometric Analyses of the 2006 MCAS High School Biology Test. Zachary R. Smith, Ronald K. Hambleton. February 4, 2008.

CEA-646 Psychometric Analyses of the 2006 MCAS High School Chemistry Test. Wendy Lam, Ronald K. Hambleton. February 9, 2008.

CEA-647 Psychometric Analyses of the 2006 MCAS High School Introductory Physics Test. Nina Deng, Ronald K. Hambleton. January 30, 2008.

CEA-648 Psychometric Analyses of the 2006 MCAS High School Technology/Engineering Test. Yue Zhao, Ronald K. Hambleton. February 17, 2008.