Embargoed until September 10, 2002

The Missouri Assessment Program

An Independent Evaluation

William D. Schafer
University of Maryland

EXECUTIVE SUMMARY

The Missouri National Education Association (MNEA) commissioned this study of the Missouri Assessment Program to evaluate the testing program's ability to contribute to Missouri's standards-based school-reform efforts. The goal is to provide recommendations that can enhance the value of the MAP within the state's school accountability system.

The Missouri Department of Elementary and Secondary Education (DESE) cooperated with MNEA and the author by supplying materials and responding to questions as they arose during the course of this study. Further, DESE has had input into the body of information that forms the basis for this evaluation. DESE's cooperation was essential to the completion of this study, and is greatly appreciated.

While this evaluation concludes that the MAP is a successful element of Missouri's student assessment program, the implementation of several recommendations could enhance its value to the state and to educators. To facilitate understanding, the recommendations from the report are grouped into four general themes below: Clarifications of Standards, Technical Considerations, Fairness and Consequences, and Reporting Results.

Clarification of Standards

Two types of standards are commonly used in an assessment program. These are content standards (what is tested) and performance standards (how well students are expected to do). These first recommendations are intended to clarify, especially for Missouri educators, precisely what Missouri's standards are.

1. For clarity of language, communicate to educators who are used to other states' terminology that Missouri's "Achievement Levels" are what others commonly call "performance standards" and that Missouri's "Performance Standards" are what others commonly call "process standards."

2. Develop assessment limits (specifications for the content that can be tested) for each sub-domain description in the state's curriculum frameworks, and describe where and how each sub-domain is assessed in the MAP through test maps (i.e., blueprints) that are useful both to educators in planning instruction as well as to test developers in creating items and other test prompts. (The supplement to the curriculum frameworks provides a good start in mathematics.)

3. Add clarity to the released scoring guides by identifying the achievement level (or scale score) that each anchor paper (example of student work) represents.

Technical Considerations

Both assessment systems (how data are generated) and accountability systems (how data are used) should conform to rigorous professional standards. Especially important in a statewide program are the quality of the data for purposes of making decisions about students, districts and schools and the effects of the program on improvement of education practices throughout the state. The purpose of this set of recommendations is to further these quality goals as well as to enhance the evidence about the degree to which they are met.

4. Replace the Terra Nova with sampling of statewide item pools as is done on the other two portions of the MAP to avoid the possibility that teachers are teaching to the specific items used on accountability measures.

5. Ensure that data from pilot and field testing are used in the revision of MAP assessments. Performance data should include frequencies and samples of student responses at all scoring-tool anchor points, especially the highest levels. Anecdotal data related to ease of administration and item validity should be collected from teachers as they administer the items.

6. Develop and disseminate the rules that are used to score MAP items so that parents, students and school personnel clearly understand how all student responses, including blanks and non-readable responses, are used in determining student test scores.

7. Conduct correlational studies relating MAP results to existing data such as teacher grades and local tests to provide evidence that scores resulting from MAP are consistent with those from other ways of measuring the same achievement.

8. Reconsider the advisability of using multiple ways to generate district performance scores for accreditation. This is a highly unusual feature of MAP that is not commonly used in other states. This feature leads to different criteria for different districts and schools, is not likely to be understood by educators or the public, and is not conducive to setting effective district and school educational goals.

9. If the state decides to continue using multiple criteria, then it should develop and include in a finalized version of the Missouri School Improvement Program (MSIP) scoring guide a justification for the different ways in which districts are awarded points for student performance. Studies to evaluate the consequences of these alternatives should be initiated as data accumulate.

10. Develop scoring and reporting systems for MAP-A (an alternate portfolio system used for the lowest-functioning students) that support accountability. These systems should assess the educational goals that students in these programs are pursuing and provide sound data that can form the basis for making outcomes-based judgments about program success.

Fairness and Consequences

Students and schools should be motivated to score well on assessments that are fair to everyone. To be fair, assessment results should have the same meaning for all. For example, they should be administered under the same conditions for everyone (i.e., be standardized). Schools should be motivated to educate toward and students should be motivated to demonstrate the highest possible levels of achievement. The recommendations in this section are oriented toward increasing motivation for students and schools, and toward ensuring that the assessments are as fair as possible.

11. Consider adding student incentives for test performance, especially for older students, to increase students' motivation to perform well on MAP assessments.

12. Monitor districts' and schools' administration of the MAP assessments to ensure that they comply with the procedures detailed in administration manuals, including procedures related to accommodations and inclusion.

13. To measure districts and schools against the same standards, establish and communicate school and district achievement targets, both for overall performance and for adequate yearly progress, based on percents of students at and above selected achievement levels for schools and districts.

14. Incorporate the Level Not Determined (LND) category into district and school accountability decisions since the mission of districts and schools is to educate all students, not just those who are tested.

Reporting Results

Assessment results are most valuable when users understand them. Fundamental to effective use is knowing what content is being assessed and how to use the information for improvement. Scores should be available on education outcomes that cover the spectrum of important goals, but at the same time are focused enough to lead to informed educational change. Finally, the scores should not be over-

interpreted beyond the quality of the data. These final recommendations are intended to enhance the impact of Missouri assessments upon schooling through the ways they are reported.

15. Monitor the success of districts with as broad an array of standardized student achievement indicators as possible—including such criteria as attendance and dropout rates—and, at the same time, minimize the impact of non-achievement factors in making decisions about district and school success.

16. Include confidence bands in reporting individual-student scale scores to convey the degree to which random factors contribute to score imprecision and to help educators avoid over-interpreting small score differences.

17. Group items in each content area tested by MAP into instructionally relevant categories and report scale-score summaries on these item categories for groups of students, such as classrooms, schools, districts and the state, so that decision making about curriculum and instructional programs can be based upon meaningful student-outcome results.

18. Discontinue reporting percent of items answered correctly for each content and process standard because differences in difficulty between the item pools have unknown impacts upon those percents.

19. Discontinue reporting process sub-scores (i.e., scores for thinking skills and processes) since they may not generalize well from one content area to another in terms of either what students are asked to do or their success in doing it.

William D. Schafer, Ed.D., is professor emeritus and affiliated professor with the Maryland Assessment Research Center for Education Success, University of Maryland-College Park. Prior to his work there, he served as state director of student assessment for the Maryland State Department of Education (1997-1999) and as a faculty member in the Department of Measurement, Statistics and Evaluation at the University of Maryland (1969-2000).

Schafer earned a doctorate in measurement, statistics and evaluation from the University of Rochester in 1969. He has authored numerous journal articles and book chapters on assessment throughout his career. He can be contacted by e-mail at ws7@umail.umd.edu.