



BOARD OF DIRECTORS

CO-CHAIRS

Louis V. Gerstner, Jr.
Chairman and CEO
IBM Corporation

Governor Tommy G. Thompson
State of Wisconsin

CO-VICE CHAIRS

Governor James B. Hunt, Jr.
State of North Carolina

John E. Pepper
Chairman of the Executive
Committee of the Board
The Procter & Gamble Company

BOARD MEMBERS

Philip M. Condit
Chairman and CEO
The Boeing Company

Governor Gray Davis
State of California

Governor John Engler
State of Michigan

George M.C. Fisher
Chairman of the Board
Eastman Kodak Company

Governor Michael O. Leavitt
State of Utah

Governor Gary Locke
State of Washington

Edward B. Rust, Jr.
Chairman and CEO
State Farm Insurance

Arthur F. Ryan
Chairman and CEO
Prudential

PRESIDENT

Robert B. Schwartz

Academic Standards and Assessments Benchmarking Evaluation for

Michigan

Presented by Achieve, Inc.

*In partnership with the Council for Basic Education and
the Learning Research and Development Center
at the University of Pittsburgh*

December 1998

Table of Contents

| | |
|--|-----|
| EXECUTIVE SUMMARY | iii |
| ACHIEVE BENCHMARKING PILOT PROJECT | 1 |
| BACKGROUND AND CONTEXT | 1 |
| BENCHMARKING TO THE BEST..... | 2 |
| THE ACHIEVE BENCHMARKING TECHNOLOGY | 3 |
| <i>The CBE Standards Benchmarking Analysis</i> | 3 |
| <i>The LRDC Standards-Assessment Alignment Analysis</i> | 5 |
| <i>The TIMSS Content Analysis of Standards and Assessments</i> | 7 |
| RESULTS FOR MICHIGAN..... | 8 |
| MAJOR FINDINGS: STANDARDS | 8 |
| <i>The English Language Arts Standards</i> | 8 |
| <i>The Mathematics Standards</i> | 10 |
| <i>The Science Standards</i> | 10 |
| <i>The Social Studies Standards</i> | 11 |
| MAJOR FINDINGS: ASSESSMENTS | 12 |
| <i>The Reading Comprehension Assessments</i> | 13 |
| <i>The Mathematics Assessments</i> | 14 |
| <i>The Science Assessments</i> | 15 |
| <i>The Social Studies Assessments</i> | 16 |
| CONCLUDING REMARKS AND RECOMMENDATIONS | 18 |

Executive Summary

In the two years since the 1996 National Education Summit, nearly every state has developed academic standards for its students and assessments to measure those standards. As states have made substantial investments in the new standards and tests, and as many states are beginning to hold students and schools accountable for performance, policymakers and the broader public want to know how their standards compare to what other states and countries expect. Achieve was created by the governors and business leaders at the Summit to help states with these issues.

Benchmarking to the Best

The Academic Standards and Assessments Benchmarking Pilot Project conducted in Michigan and North Carolina was designed to help develop a strategy for comparing a state's academic expectations against the best available national and international models. By applying the principles of benchmarking, adapted from the business world, Achieve hopes to help states answer several fundamental questions:

- How do our education standards compare with those of other high-performing states and nations?
- How rigorous are our assessments and how well do they measure the knowledge and skills set out in our standards?
- How can we strengthen our standards and assessments to improve the performance of our students?

During the Achieve benchmarking pilot, we worked with several nationally recognized organizations to design and test a comprehensive and reliable process for benchmarking a state's standards and assessments against exemplary models in each of the core academic subjects. In addition to analyzing the rigor and quality of the standards and tests, we have measured the degree of alignment between the two. During the pilot, the Council for Basic Education (CBE) conducted the standards benchmarking, while the Learning Research and Development Center (LRDC) at the University of Pittsburgh led the assessments analysis. Both organizations have extensive experience working with states and districts to design and review standards and assessments.

CBE's analysis of the quality of the Michigan standards focused on the following questions:

- How do Michigan's standards for English, mathematics, science, and social studies compare to highly regarded standards from other states and other nations?
- How clear, specific, and measurable are the Michigan standards?

LRDC's evaluation of the quality of the Michigan assessments helps address these issues:

- How well do the state assessments measure the content and skills required by the state standards?
- How challenging are the state assessments? Are the tests appropriately rigorous for the given grade level?

Results for Michigan

This report provides detailed information about the quality of Michigan's academic standards and tests. In general, the benchmarking evaluation has revealed that while the state's assessments are comprehensive and challenging, the state's standards that are published and distributed to schools and parents in the form of Curriculum Frameworks are not as clear or as articulated as they need to be to communicate the quality of the state's expectations for public school students.

- ❑ ***The Michigan academic standards are too broad and general, and crucial content found in the benchmark documents is missing in the Michigan standards.***

The state's standards sacrifice clarity and specificity for breadth and generality. In some cases, they are so broad that comparing them with standards from other states and countries is difficult. The broad language used in the Michigan standards also limits their usefulness as a vehicle for communicating with students, schools, and the public what students are expected to learn. The standards leave room for curricular choice, but do not provide enough guidance about the knowledge and skills measured by the state assessments, which are quite strong.

- ❑ ***Michigan's assessments embody a more comprehensive and demanding set of expectations for Michigan's students than might be assumed from reading the state standards alone.***

The Michigan assessments were judged by LRDC to reflect challenging and sound expectations for the relevant grade levels in each subject. This finding was generally consistent across the core academic subjects of English language arts, mathematics, science, and social studies. Overall, the assessments represent rigorous, high standards for what children should know and be able to do at key checkpoints in their schooling. These high expectations are not clearly communicated, however, through the current Curriculum Frameworks.

Concluding Remarks and Recommendations

The results of the benchmarking work conducted by CBE, LRDC, and the TIMSS Center can be best summed up the following way: the Michigan assessment system has set high expectations for students and deserves continued support, but the state has understated what those expectations are in the Curriculum Frameworks, which are the principal

documents that are shared with schools and the public. Although the tests expect students to learn demanding content in each subject, the Michigan Curriculum Frameworks are often too broad or vague to indicate what that specific content is. As a result, it is unlikely that students, their parents and teachers, and the community as a whole are fully aware of what the tests expect.

There are several reasons why this is a problem. First, the state is not providing local school districts and schools with the information they need to help their students succeed on the tests. Without adequate guidance about what the state expects students to learn at key grade levels, local districts cannot design curricula and teachers cannot shape their lessons in a way that will help prepare students to perform well on the state assessments. This scenario will raise the level of frustration among educators and could lead some to challenge or oppose the state reforms. Second, the state may be sending the public a mixed message by asking them to support and participate in the tests while being vague about what the tests are measuring. Students who score high on these tests have demonstrated achievement of valued and rigorous knowledge and skills. It is important that all who have a stake in Michigan students' achievement understand this and are able to use information from the tests to work for educational improvement.

As we see it, there are two strategies Michigan could pursue to communicate more clearly its academic standards to educators and the public. The state could revise the existing standards to better reflect the expectations on the assessments, or it could develop and distribute supplementary materials that clarify and extend the existing standards and forge a better connection with the assessments. Revising the existing standards would likely be more complicated and could send the false signal that the state is changing the direction of its reforms. On the other hand, creating new documents that would serve as a bridge between the existing standards and the assessments would continue and deepen the state's present course in education reform.

The states that have had the most success with this strategy—creating a clarifying set of materials—have made those documents the centerpiece of their public engagement campaign. In Oregon, for example, a broad set of standards similar to Michigan's Curriculum Frameworks have been superseded by a clearer, more explicit set of expectations that directly speak to the content and rigor of the assessments. Massachusetts has also produced a clarifying set of documents that bridge the standards and assessments. These new documents contain standards that better define what will be assessed on the state tests and also contain examples of test questions. Most importantly, the new sets of materials in both states are sent to schools and parents in preparation for the state tests.

We would be happy to provide examples from these and other states that have successfully bridged their standards and assessments. In the meantime, Michigan should continue to value those assessments which form the core of the Michigan Educational Assessment Program. According to our benchmarking work, these tests measure important and challenging content knowledge and skills. Students who achieve the knowledge and skills demanded by these tests have demonstrated mastery of rigorous content. The state should consider capitalizing on the quality of these assessments as Michigan moves forward with its education reform efforts.

• • •

Achieve Benchmarking Pilot Project

Background and Context

Achieve is a bipartisan, independent organization created as a result of the 1996 National Education Summit where the nation's governors and corporate leaders came together to focus their attention on the urgent need to raise academic standards and improve educational performance in America's schools. Overseen by a Board of Directors comprised of governors and corporate CEOs, Achieve's mission is to serve as a clearinghouse and resource center on education standards, testing, and accountability, and to work directly with states to support their work in these areas.

As states across the country place more emphasis on raising their standards and holding schools accountable for performance, policymakers and the broader public want to know how their expectations for students and schools compare to what other states and countries expect, especially those states and nations whose educational performance exceeds their own. In order to respond to this need, Achieve selected two national organizations with substantial experience in this field—the Council for Basic Education (CBE) and the Learning Research and Development Center at the University of Pittsburgh (LRDC)—to help develop a strategy for benchmarking a state's academic standards and assessments against the best available national and international models.

After discussions with the chief state school officers and others in their states, Governors Engler of Michigan and Hunt of North Carolina agreed to pilot Achieve's new benchmarking process in their states. Achieve, CBE and LRDC are grateful to officials in both states for their cooperation and their help in shaping this important work. We hope that the information provided in this report is useful to policymakers and others in Michigan who are committed to making the state's education system among the best in the nation and the world.

Michigan has provided model content guidelines to local districts and administered state-developed assessments in the core academic subjects for many years. In 1996, the Michigan Department of Education (MDE) developed Curriculum Frameworks in the core academic subjects; these replaced previous education standards known as the Goals and Objectives and the Model Core Curriculum that had been in place since 1985 and 1991, respectively. To measure the progress of Michigan students in meeting the standards, the state has administered assessments in reading, writing, mathematics, and science for several years to students in elementary, middle, and high school. Beginning in 1999, the state will also administer assessments in social studies in grades five, eight, and eleven. The reading, writing, and math tests are currently based on the old standards—the Goals and Objectives and the Model Core Curriculum—while the science tests given in 1998 and the social studies tests planned for next year are based on the Curriculum Frameworks. The state plans to revise the reading, writing, and math tests in the future to reflect the current Curriculum Frameworks.

The state assessments form the foundation for an accountability system focused on public reporting of individual schools' achievement results; identifying and warning schools in

trouble; providing extra financial resources to chronically under-performing schools; and applying financial penalties and taking over or closing schools that continue to fail.

While Michigan does not have a high school exit exam that all students must pass, those who achieve a high enough score on the high school assessments receive a special, state-endorsed diploma. It is worth noting that these assessments have been publicly controversial. The controversy has centered around the perceived difficulty of the assessments and their apparent incongruity with other measures that the public believes are important, such as SAT scores. In response to these criticisms, the high school tests were recently revamped; the diploma endorsement is now awarded to students whose scores fall in three of the four reporting categories, instead of only to those students who score in the top category, as was the previous practice.

Benchmarking to the Best

Benchmarking is a highly respected practice in the business world. It is an activity that looks outward to find the highest goals for performance and then measures actual business operations against those goals. Benchmarking in education is a logical extension of this business philosophy, particularly at a time when states and school districts are focused on raising standards and improving results. By benchmarking academic standards and assessments, Achieve hopes to help states answer several fundamental questions:

- Are the expectations for our students and schools high enough?
- How do our education standards compare with those of other high-performing states and nations?
- How rigorous are our assessments and how well do they measure the knowledge and skills set out in our standards?

Achieve is involved in benchmarking for another important reason: states have traditionally had limited access to quality, trustworthy information about education standards and assessments. This is partly due to the fact that the standards movement in education is relatively young. But it is also a result of the disparate nature of much of the work that has been done to date. While the standards reviews and “report cards” issued by other organizations have helped to focus national attention on the importance of quality standards, their judgments have often been in conflict and their tone has not always been constructive. States are increasingly looking for an independent, credible place to turn to for advice on these issues.

Achieve’s benchmarking efforts are not designed to grade or rank states. Instead, we are seeking to create a service that is diagnostic in nature, yielding detailed information that we hope states will find useful. In order to understand the full extent of what states expect of their students, we feel it is important not only to look at academic standards, but also at the tests states use to measure those standards. Tests are the critical link between setting standards and holding education systems accountable for achievement.

In designing a new process for benchmarking standards and assessments, we have encountered some challenges. Each of the organizations involved in this effort entered the process prepared to learn and to adapt along the way. We are confident that what we learned will not only benefit Michigan but other states as well as we continue this work in the future.

The Achieve Benchmarking Technology

We have designed a comprehensive and rigorous process for benchmarking a state's standards and assessments against exemplary models in each of the core academic subjects. The process we have developed analyzes standards and assessments as a package, rather than simply standards alone. In addition to analyzing the rigor and quality of the standards and tests, we have measured the degree of alignment between the two. During the pilot, the Council for Basic Education (CBE) conducted the standards benchmarking, while the Learning Research and Development Center (LRDC) at the University of Pittsburgh led the assessments analysis. Both organizations have extensive experience working with states and districts to design and review standards and assessments. We also took advantage of the considerable knowledge and expertise of the U.S. National Research Center for the Third International Mathematics and Science Study (TIMSS), located at Michigan State University (MSU).

The CBE Standards Benchmarking Analysis

The Council for Basic Education (CBE) designed a process for benchmarking standards to uncover the strengths and weaknesses of state standards by comparing the standards to state, national, and international "benchmark" standards that are recognized for their quality and/or for producing high student achievement. The central questions that the CBE analysis is designed to help answer are:

- How do Michigan's standards for English, mathematics, science, and social studies compare to highly regarded standards from other states and other nations?
- How clear, specific, and measurable are the Michigan standards?
- What changes can be made to strengthen the quality of the Michigan standards?

To answer these questions, CBE's staff worked with content experts and seasoned classroom teachers to develop new tools and systematically analyze the state standards. Reviewers applied sophisticated rubrics to measure how well the state standards capture the content and skills found in the benchmark standards: completely, partially, unclearly, or not at all. Similar rubrics were used to determine what kind of skills the standards expect of students, whether the state standards are clearly written, and whether the standards are measurable by local or state assessments. These criteria are explained in more detail in the accompanying findings from CBE and LRDC.

To analyze the quality of the Michigan standards in English, mathematics, science, and social studies, Achieve worked closely with CBE and LRDC to identify exemplary models of standards from other states and nations. We also consulted with officials from Michigan

and North Carolina (the other pilot state) to make sure the choices reflect their interests and needs. The benchmarks we chose for the pilot represent a range of such exemplary standards; we purposefully selected multiple examples of standards in each core subject that offer various perspectives and characteristics of good standards. We did this so that we could learn more about the different standards and so that we could provide more than one viewpoint on the Michigan standards. At the outset of the pilot, we selected four or five models in each subject, but we soon realized that it was not feasible to work with that many different documents, so we scaled back the work to 2-3 benchmark documents per subject.

Because of the increasing importance to policymakers and the public of the state-level National Assessment of Educational Progress (NAEP) assessments and the Third International Mathematics and Science Study (TIMSS) results, evidence from NAEP and TIMSS guided our selections. We used the assessment frameworks that form the basis for the NAEP assessments in reading, writing, math, science, US history, civics, and geography. While these documents provide various levels of detail about the content and skills that students need to do well on the NAEP assessments, and some are more useful than others, they are nonetheless the principal standards that communicate to the public the expectations for student achievement on these important assessments.

Likewise, we have identified several countries whose students achieve at consistently high levels in math and science throughout the years of schooling, as shown in TIMSS. We made use of the curriculum standards from one of these countries, Japan, during this pilot. Most of the other high-achieving nations' standards have not been translated and were unavailable for use during the pilot. In future work with states, we hope to incorporate more international models of quality standards.

We chose to include English language arts standards from California and New Standards (whose performance standards were developed in partnership with several states and large urban school districts) and history/social science standards from Virginia because the voluntary national standards in these subjects were not well received by states or the public. In addition, the American Federation of Teachers and the Fordham Foundation, two of the national organizations that have issued "report cards" on state standards over the last few years, have rated the California and Virginia standards very highly. As a result, there has been much interest from states and policy organizations in learning more about the standards in these states.

| Exemplary Standards Used in Benchmarking Pilot Project | | | |
|---|--------------------------------|----------------------------|---|
| <i>English language arts</i> | <i>Mathematics</i> | <i>Science</i> | <i>History/social science</i> |
| NAEP frameworks for reading and writing | NAEP framework for mathematics | NAEP framework for science | NAEP frameworks for US history, civics, and geography |
| California English language arts content standards | Japanese mathematics program | Japanese science program | Virginia history/social science standards |
| New Standards English language arts performance standards | | | |

The LRDC Standards-Assessment Alignment Analysis

The Learning Research and Development Center (LRDC) analysis of the Michigan state assessments examined the quality of those assessments and the extent to which they measure the state’s standards. Skilled judges with expertise in the content areas and test design reviewed the extent to which the state’s tests reflect the expectations set out in the Michigan standards and the extent to which scores on the state’s tests provide a full picture of students’ achievement of the standards. LRDC’s analysis helps answer the following questions:

- How well do items on the state assessments align with the expectations in the state standards?
- How challenging are the state assessments? Do the tests cover an appropriate range of difficulty for the given grade level?
- How well do the assessments as a whole represent the full range of content knowledge and skills described by the standards?

Because Michigan has developed a variety of assessments for state and local use to measure student achievement at many grade levels, LRDC consulted with senior state education officials to determine which assessments to analyze during the benchmarking pilot. LRDC focused on close analysis of the tests that are part of the state’s accountability system at key grade levels—the tests of reading comprehension and mathematics in grades

four, seven, and eleven and the assessments in science and social studies in grades five, eight, and eleven.

Each test was analyzed on an item-by-item basis, by comparing those items directly with the standards. The LRDC judges applied the following criteria to analyze whether the standards align with the state assessments:

- *Content centrality*: This criterion examines the quality of the match between the content of each test question and the content of the standard the item measures.
- *Centrality of cognitive demand*: This criterion focuses on the quality of the match between the type of cognitive demand presented by each test item and the cognitive demand described by the corresponding standard. Each item makes a certain type of cognitive demand on a student (e.g., the item requires a certain performance such as “select,” “identify,” “compare,” or “analyze”). Here, the key issue is whether there is a clear consistency between the type of performance demanded by the items and the types of performances described in the standards.
- *Challenge*: This criterion is applied to a set of items to determine whether doing well on these items requires students to master challenging subject matter. The LRDC judges consider two issues in evaluating sets of test items against the challenge criterion: the appropriate *level* of challenge and appropriate *sources* of challenge. *Appropriate level of challenge* is a judgment of whether the test items reflect a range of difficulty that is appropriate for students at the given grade level. *Appropriate sources of challenge* attempts to uncover whether the test items are difficult because of the knowledge and skills they target, or for other reasons not related to the subject matter, such as relying unfairly on students’ background knowledge or for other reasons not related to the subject matter.
- *Content representation*: This final step of the analysis measures the extent to which high scores on the set of test items mapped to each standard reflect the range and balance of content delineated in the standard. Because no one state assessment can measure the full range of knowledge and skills required by the state standards, this criterion serves as a check that the content sampled in the test is representative of what the standards describe as a whole.

The process and criteria that LRDC developed for this pilot study unpack the different factors that contribute to the alignment between standards and assessments, thereby providing a level of information about assessments typically unavailable to states. It is important to recognize that the process relies on the *judgments* of experienced content and assessment experts. The criteria and the training of judges are rigorous, but judgment is still required. LRDC provides the details of their findings and more information about the criteria in the accompanying report so that it is clear how the judgments were made.

The TIMSS Content Analysis of Standards and Assessments

To supplement and inform the work of both CBE and LRDC, we asked the TIMSS US Research Center at Michigan State University to analyze the math and science standards and assessments from Michigan using the methodology developed for the TIMSS curriculum analysis. The curriculum analysis phase of the TIMSS work began in the early 1990s, complementing the assessments administered to participating countries, and culminated in several publications documenting the academic expectations for students in over forty countries.

The US Research Center coordinated the work involved in analyzing the forty nations' curriculum documents, textbooks, and assessment items/tasks. We have been fortunate to take advantage of this comprehensive database during the pilot. The TIMSS technology systematically evaluates the content and skills that standards and assessments ask of students, and then compares these expectations with those of other nations. The Center provided Achieve with information about the math and science standards in Michigan and how they compare with the top ten highest-achieving countries in TIMSS. The results from these analyses are consistent with the analyses completed by CBE and LRDC.

Results for Michigan

Following are the major findings from the CBE and LRDC analyses of the quality of the Michigan standards and assessments. The accompanying report includes further discussion of the results that support these findings.

Major Findings: Standards

- ❑ **Michigan’s academic standards are too broad and general, and crucial content found in the benchmark documents is missing in the Michigan standards.**

Michigan’s standards sacrifice clarity and specificity for breadth and generality. In some cases, they are so broad that comparing them with standards from other states, countries, and NAEP is difficult. The broad language used in the Michigan standards also limits their usefulness as a vehicle for communicating with students, schools, and the public what students are expected to learn. The standards leave room for curricular choice, but do not provide enough guidance about the knowledge and skills measured by the state assessments, which, as noted below, are quite strong.

In each of the four core content areas, the benchmark documents take a more comprehensive approach to the disciplines than do the Michigan standards. Specifically, Michigan’s standards do not include some essential content deemed critical by the benchmarks. For example, in the area of mathematics, the Michigan standards are particularly weak in geometry when compared with the Japanese standards. Also, the Michigan reading and literature standards are missing many elements found in the reading standards recently developed for children in California. Moreover, the Michigan social studies standards do not align well with the NAEP social science frameworks. For example, in three important content areas—the development of modern America, the two world wars, and contemporary America—over 85% of the NAEP history standards have no corresponding Michigan standards.

The English Language Arts Standards

The Council for Basic Education (CBE) analyzed the content standards and benchmarks for English language arts and found that, though there are pockets of clarity and explicit content, the standards are for the most part written broadly and vaguely. When compared to the standards from California and New Standards, the Michigan English standards contain gaps in content. These findings suggest that the Michigan standards could be written more precisely and explicitly so that the state’s expectations for the study of English language arts are clear to school officials, parents, and students.

- With some key exceptions, the Michigan standards for reading and writing are generally strong when compared to the English language arts standards from California and New Standards. Michigan’s standards strongly match the California writing strategies standards, particularly for middle school and high school students, though the Michigan standards for elementary school writing are weak when compared to the California standards. Similarly, while the standards in Michigan partially or wholly match about three-fourths of the California reading comprehension standards, the Michigan standards do not compare as favorably to the California literature and vocabulary standards.
- In many cases, the Michigan standards suffer from lack of clarity. For example, CBE’s judges were unable to determine how most of the California standards for conventions (grammar, punctuation, etc.) compare to the corresponding Michigan standards for elementary school, or how the California word analysis standards for high school compare to Michigan’s standards. Other standards are characterized by excessive use of jargon, which makes it difficult for teachers, parents, and students to understand exactly what is expected. One illustration of this over-use of jargon is the following Michigan standard: “consistently use strategies to regulate the effects of variables on the communication process. An example is designing a communication environment for maximum impact on the receiver.”
- Some Michigan standards appear to partially require the content indicated for students in California, but not fully. Such examples include California’s writing application standards for late high school students, which require students to demonstrate critical skills such as writing fictional, autobiographical, and/or biographical narratives; responses to literature; reflective compositions; historical investigation reports; job applications and resumes; and multimedia presentations. The California standards take the next step and describe in detail the characteristics and features of each of the types of writing that students must produce to meet the standards. In contrast, the Michigan standards only ask students to “write fluently for multiple purposes to produce compositions, such as stories, poetry, personal narratives, editorials, research reports, persuasive essays, resumes, and memos.”
- On the whole, the Michigan expectations are reasonably strong in reading, writing, and literature when compared to the New Standards English language arts standards, although in many cases the New Standards expectations are more explicit and precise. However, for the majority of the New Standards expectations for what elementary, middle and high school students should know about conventions, it is unclear whether the corresponding Michigan standards contain similar content because the Michigan standards are vague. Michigan’s standards for speaking, listening, and viewing are also weak: only 15% of the standards match the corresponding New Standards expectations.

The Mathematics Standards

The Council for Basic Education (CBE) compared the Michigan mathematics standards to the standards for Japanese students and to the framework for the National Assessment of Educational Progress (NAEP). Overall, the Michigan standards are missing significant content and skills and do not align strongly with either the NAEP or Japanese standards.

- When benchmarked against the Japanese expectations, the Michigan standards receive an overall poor rating. Less than 10% of the standards for Japanese elementary school students are matched by Michigan's standards and only a quarter of the Japanese high school standards are found in the Michigan standards. On the other hand, at the middle school level, the Michigan standards are moderately aligned to the Japanese standards. Over two-thirds of the Japanese standards for eighth grade numbers and quantitative relations match the Michigan standards.
- Evidence from the TIMSS assessments indicates that American students have a weak command of geometry, measurement, and algebra, while students in high-achieving countries such as Japan and Singapore perform well in these areas. The CBE analysis and the TIMSS Center analysis indicate that the Michigan standards could be strengthened in these areas, particularly in middle and high school.
- All Japanese students study algebra and geometry throughout junior high school (grades 7-9) and are required to take Mathematics I in tenth grade. In this course, students learn advanced topics in algebra, trigonometry, statistics, and probability that may be reserved for college-bound students in Michigan. Moreover, CBE's analysis indicates that none of the content and skills related to geometry and advanced algebra that are required in the Japanese standards for middle and high school are included in the Michigan standards.
- Michigan's standards compare slightly more favorably to the expectations laid out in the NAEP framework, but a significant gap exists between the two. On average, about half of the standards in NAEP are included in Michigan. The alignment between Michigan and NAEP is strongest for eighth grade data standards and weakest in the areas of number and measurement throughout the grades.
- According to CBE, some important mathematics concepts are not adequately represented in the Michigan standards. These include rounding, conics, and the Pythagorean Theorem.

The Science Standards

CBE compared the Michigan science standards to the standards for Japanese students in grades three through nine and to the framework for the National Assessment of Educational Progress (NAEP). Overall, the material covered in the Japanese standards is only addressed marginally in the Michigan standards. The content outlined in the NAEP framework is more thoroughly, but not completely, represented in Michigan.

- The Michigan standards are not well aligned to the Japanese expectations for science. Almost half of the Japanese standards are completely excluded from the Michigan framework and less than one-quarter of the Japanese standards are matched by Michigan expectations. Some of this poor alignment can be attributed to broad or vague language used in Michigan, but more often the Michigan standards are simply missing the content and skills expected of schoolchildren in Japan.
- There is a slightly stronger match between the standards from NAEP and Michigan than exists between the Japanese and Michigan science standards. Interestingly, Michigan’s expectations are more consistent with NAEP’s standards for elementary school than for high school. For example, over three-fourths of NAEP’s fourth grade standards for earth science are well-aligned to the Michigan standards. As with the comparison to the Japanese standards, the gaps in Michigan’s standards relative to NAEP are more often the result of missing content than vague or broad language.
- CBE noted that the format of the standards is difficult to understand. In each standard, the lead sentence provides the big idea of the standard. Two separate lists—Key Concepts and Real-World Contexts—follow the lead sentence. The problem with this format is that these lists are much too general to provide enough information about what is required of student learning. Moreover, the language of the Michigan standards may be overly dependent upon words such as “describe” and “explain,” which could inadvertently send the message that all students need to be able to do is memorize content and learn it superficially.

The Social Studies Standards

The lack of consistency in the organization and structure of the benchmark standards in history/social studies made the comparative work in this subject particularly challenging and required choices to be made. NAEP has produced separate frameworks for US history, civics, and geography. Because the expectations contained in these frameworks are extensive, the CBE judges compared the Michigan standards for kindergarten through eighth grade against the NAEP expectations in these subjects for eighth grade. Similarly, the elementary and middle school history/social science standards from Virginia were compared to the Michigan standards for kindergarten through eighth grade. All of Michigan’s standards, including those for high school, were compared to Virginia’s standards for world history in grades 8 and 9 and US history in grade 11.

Overall, the Council for Basic Education (CBE) found that the Michigan standards are missing a considerable amount of the content and skills set out in the NAEP frameworks and Virginia standards. This is somewhat understandable in the case of NAEP, given the voluminous amount of material covered by these assessments. It would be difficult, and perhaps inappropriate, for any state to cover it all. Still, there are numerous essential topics and content areas missing from the Michigan standards that can and should be addressed.

- There is very little match between the content in the NAEP frameworks for US history, civics, and geography and the content of the Michigan social studies standards. The most glaring difference is in the area of civics. Only 3% of the NAEP civics standards are completely addressed in the Michigan social studies standards. The Michigan social studies standards align poorly to the NAEP US history and geography standards as well. Over fifty percent of the history and geography standards that NAEP identifies as important for students to learn are excluded in Michigan.
- In some cases, the low correlation between NAEP and Michigan can be attributed in part to the vague or broad language used in the Michigan standards. The Michigan standards are not specific about content; instead, they offer broad statements about the social studies skills that students should acquire, without indicating the events, eras, or ideas students need to learn and examine in order to develop and apply those historical skills.
- Likewise, the Virginia history/social science standards are more comprehensive and challenging than the Michigan standards. There is very little correlation between the content included in the Virginia standards and the Michigan social studies expectations. The Virginia content area with the highest percentage of match to Michigan is geography, and even so the Michigan standards only match 33% of the corresponding Virginia standards. Most of this weak alignment is caused by the absence of historical content in the Michigan standards.
- CBE also found that there is a noticeable lack of world history in the Michigan standards. World history only seems to appear in the context of studying world geography.

Major Findings: Assessments

- ❑ **Michigan’s assessments embody a more comprehensive and demanding set of expectations for Michigan’s students than might be assumed from reading the state standards alone.**

The Michigan assessments were judged by the Learning Research and Development Center (LRDC) to reflect challenging and sound expectations for the relevant grade levels in each subject. This finding was generally consistent across the core academic subjects of reading, mathematics, science, and social studies. There are some qualifications related to particular areas of the tests that are identified in the accompanying findings from CBE and LRDC. Overall, however, the assessments represent rigorous, high standards for what children should know and be able to do at key checkpoints in their schooling.

According to senior education officials, it is practice in Michigan to give tests that are based on older standards until school districts have had sufficient time to implement the new standards. For this reason, the current assessments in reading and mathematics are based on previous standards—known as the Essential Goals and Objectives, in place since

1985, and the Michigan Model Core Curriculum, adopted in 1991—rather than the Curriculum Frameworks, which were approved in 1996. Because of the state’s cycle for revising the state assessments, the current science and social studies tests are based on the new Curriculum Frameworks, while the math and reading tests are based on the older standards. Thus, the tests that were studied by LRDC are based on three different sets of standards. In each subject discussed below, it is clearly noted whether the tests were built on the former standards or the current Michigan standards.

The Reading Comprehension Assessments

LRDC analyzed the reading tests given to all students in grades four, seven, and eleven. The tests are based on Michigan’s Essential Goals and Objectives, which precede the Curriculum Framework. The Framework was developed in 1996 and forms the basis for the CBE analysis of the Michigan state standards. In the future, the state plans to align the reading tests to the new Curriculum Framework.

LRDC judged the reading tests to be appropriately challenging and demanding in terms of the knowledge and skills they require. The tests for grades four and seven rely exclusively on multiple-choice questions. The high school test also includes an open-ended essay response to the reading passages.

- The three reading tests align strongly with the content and skills laid out in the Goals and Objectives. LRDC judged the tests to be challenging and fair assessments of reading at the elementary, middle, and high school levels. Each test represents a good range and balance of the various elements of the Objectives that can be fairly measured by an on-demand test in the selected format. The tests also place appropriate emphasis on those elements of reading of greatest concern at each level.
- In a multiple-choice format reading test, the reading passages selected determine how difficult the test is at least as much as the questions related to these passages. The selections chosen for the tests in grades four and seven are appropriate and engaging for these grade levels. One of the informational texts appears to be drawn directly from a social studies text, which is a good source of authentic or “real-life” reading passages for students at these levels. Moreover, none of the questions on these tests presents a “trick” for the test-taker; however, as should be the case, some questions are more demanding than others.
- The reading passages included in the high school test provide a variety of settings and types of texts and, according to LRDC, represent an appropriate range of reading materials found in high school English or social studies classrooms. The questions related to these passages are demanding yet appropriate for high school students.
- According to LRDC, another aspect of the reading passages on the tests deserves mention. Each test includes 2-3 passages, whereas many reading assessments are built on a large number of short reading passages in an attempt to sample as wide a

range of reading types as possible. By providing ample time, engaging and appropriate readings, and a limited number of passages, the Michigan tests provide a better approximation of real reading situations and, thus, present a more valid assessment of students' reading abilities.

The Mathematics Assessments

LRDC analyzed the mathematics tests for grades four and seven and judged them to be reasonably demanding assessments of the knowledge and skills required by the Essential Goals and Objectives. In consultation with officials from the Michigan Department of Education, LRDC decided not to conduct the alignment analysis on the high school math assessment. The test blueprint used to develop the assessment provides limited information about the relationship between the standards and the tests. Thus, the alignment analysis would have been impossible to complete without requiring LRDC to make undue levels of interpretation. In the future, the state plans to align the math tests (including the high school assessment) to the new Curriculum Framework. Again, the CBE analysis points out several significant content gaps between the standards in the Curriculum Framework, which as mentioned earlier, contain significant holes when compared to the Japanese and NAEP standards.

- On the whole, the tests for grades four and seven are aligned moderately well with the content and skills found in the Goals and Objectives. Also, generally speaking, both tests are appropriately challenging for students at these levels.
- On the elementary school assessment, the questions measuring the standards for geometry and problem solving and reasoning are less well aligned to the Goals and Objectives than the questions covering other content areas. The questions about fractions and decimals are relatively strongly aligned in terms of the content and skills they target.
- Both tests represent a good range and balance among the content areas of mathematics identified as important by the standards. More than one-third of the fourth grade test focuses on whole numbers and the seventh grade test has a relatively high emphasis on questions about fractions and decimals, both of which are appropriate for these grade levels.
- The grade seven assessment under-represents number properties and problem solving applications related to measurement. This is noteworthy for two reasons: first, results from the Third International Mathematics and Science Study indicate that American middle school students score poorly in measurement, while students in the highest-achieving countries have mastered this important content area by this age. Secondly, the Goals and Objectives on which this test was built only include measurement topics as focal points in grades 4-6, so the seventh grade test is the optimal and only time to assess how well students understand and can apply measurement concepts.

- In some cases, it is difficult to determine whether the assessments measure the range of mathematical content required of students. This is particularly true in geometry. Twelve of the 14 geometry standards are identical across the three grade levels and these standards are expressed at a more general level than the standards related to other areas of math.
- The use of an entirely multiple-choice format for the mathematics tests for grades four and seven limits the range of challenge that may be presented by the test. By definition, an item in a multiple-choice format includes one answer choice that is correct; sometimes the easiest way to find the answer is to follow a process of elimination rather than solve the problem presented in the item. It is important to note, however, that the multiple-choice format does not preclude questions that require more complex thinking. The limitation noted in this case by LRDC is most evident among the questions on both tests related to problem solving. This limitation restricts the potential for the test to target the more challenging types of performances described in the Goals and Objectives.

The Science Assessments

LRDC analyzed the science assessments for grades five, eight, and eleven; unlike the reading and mathematics assessments, these tests were developed using the Curriculum Framework written in 1996. LRDC's analysts evaluated the sections on these three assessments that include both multiple-choice and constructed-response test questions. According to LRDC, the tests are generally well aligned to the science standards, which were judged weak in CBE's analysis. However, throughout the three tests, certain key concepts and skills are not assessed sufficiently.

- Overall, about two-thirds of the items on the fifth grade and high school tests and about 90% of the items on the eighth grade test align strongly with the standards in terms of the content that is targeted and the performances that the standards require.
- The accompanying report from LRDC provide explicit information about how well the science assessments measure the breadth and depth of the scientific knowledge and skills discussed in the standards. In general, the three tests do not cover the full range of content knowledge and scientific skills that are included in the standards. While no assessment can comprehensively cover all content, some important areas of scientific knowledge are consistently under-represented on the tests.
- On the fifth grade test, some questions provide a balanced and reasonable representation of the content that is measured, but other content areas, notably physical science (energy and static electricity), life science (ecosystems), and technology are under-represented on the assessment. On the 8th grade test, important concepts that are neglected include basic cell biology, relationships (e.g, electricity and magnetism, populations, food webs, cycling of matter), and applications of scientific concepts to daily life (i.e., the relative benefits and risks of technology).

- The high school assessment similarly neglects important scientific knowledge, including content related to the use of technology; basic cell biology, evolution, and heredity; and energy (conversion in living organisms, machines, chemical and physical changes). The omission of these concepts throughout the three assessments—particularly of cell biology and energy—could be perceived as a serious weakness of the tests, given both the fundamental role of cell biology as a building block of life science and the poor showing of American students in the area of energy in TIMSS.
- Overall, the three tests were judged to reflect an appropriate range of difficulty for students at these levels. However, some questions on the high school test related to important subject matter tend toward the low end of difficulty for students at this level. The questions that are not as challenging as the standards indicate they should be are related to cell biology; changes in matter; and key concepts in earth and space science including atmosphere, weather, and the solar system, galaxy, and universe.

The Social Studies Assessments

LRDC analyzed the recently-developed social studies assessments for grades five, eight, and eleven that are based on the Curriculum Framework and will be administered for the first time in 1999. Overall, the three assessments align moderately to strongly with the content and performances expected by the standards.

- Each test measures students' knowledge of history, civics, economics, and geography, and generally the test questions align moderately to strongly with the corresponding standards. However, there are some content areas that are not addressed as strongly as others and oftentimes the test questions are less challenging than the standards would indicate they should be.
- On the elementary school test, all questions align either moderately or strongly with the content in the standards and the performance expectations indicated by the standards. All of the questions related to public discourse and decision-making and inquiry are aligned strongly with the standards, while questions related to historical perspective, geographic perspective, and economic perspective are moderately aligned to the content and skills found in the standards.
- The test for eighth grade reflects a similar pattern. The questions related to historical perspective and geographical perspective align well with the content found in the standards, but the items usually ask for performances that are less cognitively demanding than the standards require. Moreover, most of the test items related to civic perspective align moderately with the standards' content and performance expectations. The questions related to economic perspective generally align only moderately or weakly with the relevant content and skills from the standards. For example, one item asks for identification of positive and negative consequences of deregulation of an industry, whereas the related standard requires

explanation and demonstration of how businesses confront scarcity and choice when organizing.

- On the high school assessment, the questions related to historical perspective and economic perspective generally reflect a stronger alignment with the content of the relevant standards than with the cognitive demand found in the standards. Questions focused on geographical perspective and civic perspective align moderately with the standards' content and performance expectations.
- The majority of test questions on the three assessments fall within the range of difficulty that might be expected from reading the standards. However, there is some evidence that some questions on the elementary and middle school tests present a lower level of challenge than the standards require. The eighth grade test includes several questions that ask students to recall specific pieces of information or to identify information from data presented in a graphic presentation, rather than requiring them to engage in more complex processes of analysis and evaluation, which are required by the standards. Generally, the middle school test includes few questions that require students to draw comparisons and contrasts, such as between differing perspectives or historical periods, or to interpret information with consideration of existing/changing conditions or issues.
- The three tests provide a reasonably comprehensive and balanced representation of the content demanded by the related standards, with two exceptions. On the elementary school assessment, the sets of questions related to geographical perspective, civic perspective, and economic perspective tend to over-represent small, isolated elements of those areas of content, rather than reflecting the larger ideas that frame the content of the standards. The middle school test questions dealing with the standards for civic perspective and economic perspective provide a good representation of certain elements of the standards, but tend to overlook other elements of the standards that are also important. For example, questions related to economic perspective focus on national and international issues but do not include questions related to individual and household economic decisions.

Concluding Remarks and Recommendations

It is important to preface our concluding thoughts by once again thanking Governor Engler, Superintendent Ellis, and other senior education officials in Michigan for offering to serve as a pilot site for Achieve's initial benchmarking work. The work was intensive and Achieve, the Council for Basic Education, and the Learning Research and Development Center are very appreciative of the support we received from state officials. We also admire Michigan's willingness to examine critically their work on standards and assessments and to think constructively about how that work can be continually improved. This is certainly indicative of the leadership the state has shown in education reform.

The results of the benchmarking work conducted by CBE, LRDC, and the TIMSS Center can be best summed up the following way: the Michigan assessment system has set high expectations for students and deserves continued support, but the state has understated what those expectations are in the Curriculum Frameworks, which are the principal documents that are shared with schools and the public. Although the tests expect students to learn demanding content in each subject, the Michigan Curriculum Frameworks are often too broad or vague to indicate what that specific content is. As a result, it is unlikely that students, their parents and teachers, and the community as a whole are fully aware of what the tests expect.

There are several reasons why this is a problem. First, the state is not providing local school districts and schools with the information they need to help their students succeed on the tests. Without adequate guidance about what the state expects students to learn at key grade levels, local districts cannot design curricula and teachers cannot shape their lessons in a way that will help prepare students to perform well on the state assessments. This scenario will raise the level of frustration among educators and could lead some to challenge or oppose the state reforms. Second, the state may be sending the public a mixed message by asking them to support and participate in the tests while being vague about what the tests are measuring. Students who score high on these tests have demonstrated achievement of valued and rigorous knowledge and skills. It is important that all who have a stake in Michigan students' achievement understand this and are able to use information from the tests to work for educational improvement.

As we see it, there are two strategies Michigan could pursue to communicate more clearly its academic standards to educators and the public. The state could revise the existing standards to better reflect the expectations on the assessments, or it could develop and distribute supplementary materials that clarify and extend the existing standards and forge a better connection with the assessments. Revising the existing standards would likely be more complicated and could send the false signal that the state is changing the direction of its reforms. On the other hand, creating new documents that would serve as a bridge between the existing standards and the assessments would be more of a continuation of the present course.

The states that have had the most success with this strategy—creating a clarifying set of materials—have made those documents the centerpiece of their public engagement

campaign. In Oregon, for example, a broad set of standards similar to Michigan's Curriculum Frameworks have been superseded by a clearer, more explicit set of expectations that directly speak to the content and rigor of the assessments. Massachusetts has also produced a clarifying set of documents that bridge the standards and assessments. These new documents contain standards that better define what will be assessed on the state tests and also contain examples of test questions. Most importantly, the new sets of materials in both states are sent to schools and parents in preparation for the state tests.

We would be happy to provide examples from these and other states that have successfully bridged their standards and assessments. In the meantime, Michigan should continue to value those assessments which form the core of the Michigan Educational Assessment Program. According to our benchmarking work, these tests measure important and challenging content knowledge and skills. Students who achieve the knowledge and skills demanded by these tests have demonstrated mastery of rigorous content. The state should consider capitalizing on the quality of these assessments as Michigan moves forward with its education reform efforts.

• • •

Our goal with this pilot project was to develop a new set of procedures for analyzing standards and assessments; to apply those procedures in two states; and to provide those states with compelling, constructive information that can be used to strengthen their expectations for students and schools. We hope the information we have provided here and in the accompanying findings from CBE and LRDC is helpful to Michigan as it continues to work toward a higher performing and a more accountable education system.