STATE OF MICHIGAN
DEPARTMENT OF EDUCATION
LANSING

IENNIFER M. GRANHOLM
GOVERNOR

MICHIGAN Department of Education

THOMAS D. WATKINS, JR.
SUPERINTENDENT OF
PUBLIC INSTRUCTION

TO: Tom Watkins, Superintendent

FROM Jeremy M. Hughes, Ph. D.
Chief Academic Officer/Deputy Superintendent

DATE: April 8, 2003

SUBJ: *EDUCATION YES!* STATE BOARD OF EDUCATION ISSUES

We are scheduled to make a progress report to the State Board at its committee-of-the-whole meeting on April 17, 2003. As background for that discussion, I am:

Attaching the latest draft of the report of the Accreditation Advisory Committee (AAC).

- Presenting in this memo the issues that need to be presented to and discussed by the Board.

Grading "Growth"

Pages 10-12 of the AAC's report address this issue. The AAC previously recommended that "Growth" between elementary and middle school not be graded until such time as the new MEAP tests, grades 3-8, are in place. The AAC stated that because the 7th grade English Language Arts test, for instance, is not equated with the 4th grade ELA test, there are statistical problems involved in comparing the results on one with the results on the other.

The AAC provided an "interim solution," however (in Appendix A of its report), to be used in the event the State Board wished to proceed anyway to grade elementary to middle school Growth. The State Board did indicate a wish to proceed.

The AAC also pointed out (p. 11) that grading Growth at the high school level posed a more serious measurement problem and recommended that, even if the State Board wished to proceed to grade Growth between elementary and middle school, it not do so for Growth between middle and high school.

608 WEST ALLEGAN STREET • P.O. BOX 30008 • LANSING, MICHIGAN 48909
www.michigan.gov/mde • (517) 373-3324

Department staff have been at work on implementing the "interim solution" recommended by the AAC for Growth between elementary and middle school. In the course of planning, the issues cautioned by the AAC have become apparent.

It will therefore be the recommendation of staff that the decision of the State Board to proceed with the "interim solution" be reconsidered, and that grading Growth between elementary and middle school as well as between middle and high school, be delayed until the new grade-level tests required by NCLB are in place and until the options for calculating Growth between middle and high school can be analyzed further.

MEAP "Status" - New vs. Old MEAP Tests

Pages 12-14 of the AAC report address this issue. The AAC, early in its deliberations, recommended that grades for MEAP Status, Change, and Growth be based on three-year averages. Doing this would "soften" the impact for a school of MEAP scores that one year might be uncharacteristically low, or uncharacteristically high.

The problem, however, lies in the fact that:

- The English Language Arts (ELA) test is new, and was administered for the first time this year, 2002-03.
- The new ELA test will have scale scores for reading and for writing, but will not report a combined "ELA" scale score.
- The Mathematics and Science tests were new in 2001-02, and were administered for the second time this year, 2002-03.

If the three-year-average rule was adhered to, this would mean that this year's scores on the new ELA test would be averaged with two previous years' scores on the former Reading test. The new and old tests are not equated. Simple "averaging" will not be appropriate.

Similarly, for Mathematics and Science, the 2001-02 and 2002-03 scores on the new tests would be averaged with one year (2000-01) scores from the old tests. The new Mathematics and Science tests are not equated with their predecessor tests.

The AAC is recommending:

- For English Language Arts – do not use the one-year results from the new test on this year's report card. Instead, compile the three-year average using the three most recent years with the former Reading test. For next year (2003-04) calculate the report card based on two, not three, years of scores, using what would next year be two years of scores from the new tests.

For Mathematics and Science – calculate report card grades this year based on only two years of scores, not three, using the scores from the new tests from 2001-02 and 2002-03. For next year (2003-04), we would have a third year of scores from the new tests and the three-year average would be able to be calculated.

These issues are not applicable to the Social Studies test because that test has not changed.

It will be the recommendation of staff that the AAC's proposal be accepted. The approval of the State Board of Education is thought to be necessary because the above proposal departs, in the cases mentioned, from the Board's direction that three-year averages be used to compute report card grades.

On the issue of the English Language Arts test, staff recommends that for the purpose of calculating achievement status, the separate reading and writing scores be used with the weighting as planned for the English Language Arts test.

MEAP "Change" New vs. Old MEAP Tests

For achievement change, the AAC recommends that the School Change cut-scores and grades that ideally should be calculated using three 3-year averages, or five years of comparable data. The AAC recommendation on this issue is on pages 13 and 14 of the report. The committee recommends using data from the old test until three years of data from the new test are available. This approach will require use of the new tests until the third year that it is administered.

Staff have developed an alternative that moves up the use of the new tests in the change calculation. This approach would average two slopes computed from three years of the old test and two years of the new test. This would be followed by averaging two years of the old test with three years of the new test, then four years of the new test. This approach uses more years of data, resulting in a less volatile calculation.

It will be the recommendation of staff that the AAC's proposal be accepted with the modification of averaging the slopes calculated from three years under the old test and two years of the new test.

Note: the following table illustrates the recommendations of the AAC regarding the treatment of old vs. new MEAP tests:

MEAP results used in calculating and reporting Education YES! cut-scores & grades

Subject Area	In 2003	In 2004	In 2005	In 2006	In 2007
Reading					
Status	00-01-02	03-04	03-04-05	04-05-06	05-06-07
Change	98-99-00-01-02	98-99-00-01-02	03-04-05	03-04-05-06	03-04-05-06-07
Growth	01-02	03-04	04-05	05-06	06-07
Math & Science					
Status	02-03	02-03-04	03-04-05	04-05-06	05-06-07
Change	98-99-00-01-02	02-03-04	02-03-04-05	02-03-04-05-06	03-04-05-06-07

School Performance Indicators

Pages 17-18 of the AAC's report address this issue. The AAC has repeatedly expressed its reservations about the grading of the indicators, citing the following reasons:

- The questions about the indicators are more "process based" than "data based," that is, schools were asked if they had policies or practices in place vs. asking them for data to indicate whether those policies or practices were effective.

- More than 50% of the more-than-3,000 schools responding gave themselves top marks on the majority of the indicators. With so many schools bunched up on one end of the continuum, the AAC felt it was not able, statistically, to recommend A, B, C etc. cut scores.

In its draft final report (p. 18), the AAC recalled its earlier recommendation that the School Performance Indicators not be included in the report card "until such time as improved measures are available." In the event, however, that the State Board of Education wished to proceed anyway, an "interim solution" was offered, as described on page 18.

The primary element of the AAC's recommendation is that, if included in the report card, the 33-point weighting originally assigned to the indicators be reduced to an 11-point weighting. These 11 points would then be combined with the 67 points assigned to the MEAP grades, resulting in a report card based on 78 vs. 100 points, as originally mandated by the State Board.

Since the final meeting of the AAC on March 10, 2003, department staff have been at work analyzing the data from the 3,000+ schools. Our preliminary work suggests that:

- We can devise a system to grade the indicators, acknowledging that a majority of schools will receive relatively high grades for the indicators on this first report card.

- The "grading" of the school performance indicators should result in a single grade for the 11 indicators, rather than three separate grades for engagement, instructional quality, and learning opportunities.

- By assigning one grade to all eleven of the indicators, we can maintain the 33-point weighting originally assigned by the State Board.

It will therefore be our recommendation to the State Board that the School Performance Indicators be included in the first report card, and that the 33-point weighting be retained with a single grade assigned for all eleven of the indicators.

Initial Year of Statewide Implementation

Pages 22-25 of the AAC report address this issue. The AAC felt strongly that the first year issuance of school report cards is only a beginning step in the refinement of a high quality school accreditation system for Michigan. To that end, the AAC listed seven recommendations for further study, revision, refinement, and research.

Although action by the State Board of Education on these issues would be premature – and perhaps in the end not even necessary – department staff feel that the State Board should be aware of the actions, challenges, and resources the AAC feels are necessary if the *Education YES!* system is to be viable and earn the acceptance and support of the school and pyschometric communities.

It is the intention of department staff to briefly review the AAC's "first year of implementation" recommendations at the committee-of-the-whole meeting.

FINAL REPORT AND RECOMMENDATIONS OF THE MICHIGAN ACCREDITATION ADVISORY COMMITTEE TO THE STATE BOARD OF EDUCATION

Introduction

The recommendations set forth in the following pages are offered to the State Board of Education in fulfillment of the charge set before us on our appointment as the Michigan Accreditation Advisory Committee. The members of our committee are five in number and include:

Philip Kearney	University of Michigan – Ann Arbor
Sharon Johnson Lewis	Council of the Great City Schools – Washington, DC
Lawrence Lezotte	Effective Schools Products, Ltd - Okemos
Mark Reckase	Michigan State University – East Lansing
Edward Roeber	Measured Progress – Dover, NH

As noted in Superintendent of Public Instruction Thomas Watkins' memorandum to the State Board of Education dated April 29, 2002, each of us brought to the work of the committee some considerable expertise and experience in accountability, measurement, school improvement, and accreditation systems.

In that same memorandum, Superintendent Watkins laid out the committee's charge, namely to develop recommendations in three areas:

- Initial Distribution of Schools in Grade Categories
- Measuring School Performance Indicators
- Alignment of *Education YES!* with Federal Legislation

In fulfilling our charge, the committee met as a group on twelve separate occasions:

May 8-9, 2002	Crowne Plaza, Ann Arbor & Monroe ISD, Monroe
June 10-11, 2002	Washtenaw ISD, Ann Arbor
July 15-16, 2002	Washtenaw ISD, Ann Arbor
August 1-2, 2002	State Library, Lansing
September 9-10, 2002	Washtenaw ISD, Ann Arbor
October 10-11, 2002	Washtenaw ISD, Ann Arbor
October 31-November 1, 2002	Washtenaw ISD, Ann Arbor
November 18-19, 2002	Washtenaw ISD, Ann Arbor
December 9-10, 2002	Washtenaw ISD, Ann Arbor

Note: Text in italics indicates material that is not yet available, e.g. high school cut scores, for which MEAP data are not yet available.

January 21-22, 2003	Washtenaw ISD, Ann Arbor
February 14, 2003	Washtenaw ISD, Ann Arbor
March 10, 2003	Washtenaw ISD, Ann Arbor

The committee met by telephone conference call on October 21,2002.

In addition, one or more members of the Committee provided interim reports to the State Board of Education on the following occasions:

September 14, 2002	Philip Kearney	State Board regular meeting
October 24, 2002	Philip Kearney	State Board regular meeting
November 14, 2002	Edward Roeber	State Board regular meeting
December 9, 2002	Full Committee	State Board committee of the whole meeting
December 12, 2002	Philip Kearney Larry Lezotte	State Board regular meeting

The Committee's Recommendations

As called for in its charge, the Committee has developed and offers recommendations in the following three areas:

- Criteria for Assigning Scores/Grades
- Measuring School Performance Indicators
- Alignment of *Education YES!* with Federal Legislation

In addition, the committee offers recommendations in a fourth area:

The Initial Year of Statewide Implementation and Further Development of the System

The recommendations in all four areas are set forth in the following pages.

I. Criteria for Assigning Scores/Grades

The Committee recommends the following cut scores/letter grades for School Status, School Change, and Student Growth.

For School Status

The School Status score/grade is an index score that reflects a school's success vis-à-vis academic achievement, taking into account the school's success in achieving academic proficiency for all students. The method of calculating the individual index scores for Michigan schools is set forth in detail in Appendix A.

The committee recommends the following assignments of cut scores/grades for School Status. The distributions of cut scores/grades for School Status are also provided:

Elementary Reading Status				
Grade	Index Score Range	Number of Schools	Percent of Schools	Cumulative Percent of Schools
A	Above 311	210	13.1%	13.1%
B	307 – 310	341	21.3%	34.4%
C	298 – 306	647	40.4%	74.8%
D	294 – 297	160	10.0%	84.8%
F	Below 293	244	15.2%	100.0%

The index score range is expressed in terms of the weighted achievement index recommended by the Accreditation Advisory Committee.

Elementary Mathematics Status				
Grade	Index Score Range	Number of Schools	Percent of Schools	Cumulative Percent of Schools
A	Above 537	354	22.1%	22.1%
B	528 – 536	409	25.5%	47.6%
C	513 – 527	590	36.8%	84.4%
D	509 – 512	111	6.9%	91.3%
F	Below 508	139	8.7%	100.0%

The index score range is expressed in terms of the weighted achievement index recommended by the Accreditation Advisory Committee.

Middle School Reading Status				
Grade	Index Score Range	Number of Schools	Percent of Schools	Cumulative Percent of Schools
A	Above 305	69	10.2%	10.2%
B	298 – 304	205	30.3%	40.5%
C	295 – 297	179	26.5%	67.0%
D	288 – 294	134	19.9%	86.9%
F	Below 289	88	13.1%	100.0%

The score range is expressed in terms of the weighted achievement index recommended by the Accreditation Advisory Committee.

Middle School Mathematics Status				
Grade	Index Score Range	Number of Schools	Percent of Schools	Cumulative Percent of Schools
A	Above 527	176	26.1%	26.1%
B	523 – 526	85	12.5%	38.6%
C	512 – 522	268	39.6%	78.2%
D	502 – 511	100	14.8%	93.0%
F	Below 501	47	6.9%	100.0%

The index score range is expressed in terms of the weighted achievement index recommended by the Accreditation Advisory Committee.

Middle School Science Status				
Grade	Index Score Range	Number of Schools	Percent of Schools	Cumulative Percent of Schools
A	Above 399	136	20.6%	20.6%
B	387 – 398	134	20.3%	40.9%
C	373 – 386	156	23.5%	64.4%
D	341 – 372	167	25.2%	89.6%
F	Below 342	69	10.4%	100.0%

The index score range is expressed in terms of the weighted achievement index recommended by the Accreditation Advisory Committee.

Middle School Social Studies Status				
Grade	Index Score Range	Number of Schools	Percent of Schools	Cumulative Percent of Schools
A	Above 517	143	21.8%	21.8%
B	512 – 516	72	10.9%	32.7%
C	503 – 511	166	25.3%	58.0%
D	484 – 502	197	30.0%	88.0%
F	Below 483	80	12.1%	100.0%

The index score range is expressed in terms of the weighted achievement index recommended by the Accreditation Advisory Committee.

(On December 12, 2002, the State Board of Education adopted the above recommendations on cut-scores and grades.)

Grade 11 Reading

The recommendation regarding cut scores for achievement status in reading at the high school level will be inserted at such time as high school MEAP data are available in a format that allows calculation of the weighted achievement index recommended by the Accreditation Advisory Committee.

Grade 11 Writing

The recommendation regarding cut scores for achievement status in writing at the high school level will be inserted at such time as high school MEAP data are available in a format that allows calculation of the weighted achievement index recommended by the Accreditation Advisory Committee.

Grade 11 Mathematics

The recommendation regarding cut scores for achievement status in mathematics at the high school level will be inserted at such time as high school MEAP data are available in a format that allows calculation of the weighted achievement index recommended by the Accreditation Advisory Committee.

Grade 11 Science

The recommendation regarding cut scores for achievement status in science at the high school level will be inserted at such time as high school MEAP data are available in a format that allows calculation of the weighted achievement index recommended by the Accreditation Advisory Committee.

Grade 11 Social Studies

The recommendation regarding cut scores for achievement status in social studies at the high school level will be inserted at such time as high school MEAP data are available in a format that allows calculation of the weighted achievement index recommended by the Accreditation Advisory Committee.

NOTE: The committee recommends that, at the high school level, the State Board of Education consider using the MEAP scores in the place of the percent who qualify for the Michigan Merit Award (which currently can be earned via MEAP, ACT, SAT or Work Keys). The committee makes this recommendation because it anticipates that the *No Child Left Behind Act (NCLB)* requires states to assess all students at one high school grade. Since NCLB Title I uses

mathematics and reading/language arts scores to gauge Adequate Yearly Progress (AYP), we recommend that the MEAP tests be used to accredit/grade Michigan high schools. This may necessitate change in the MEAP testing policies.

The percent qualifying for the Michigan Merit Award could still be reported, but not included in the school's grades.

(On November 14, 2002, the State Board of Education adopted the above recommendation and modified Education YES! to reflect this change.)

For School Change

A target slope will be calculated for each school based on a school's performance against its own target slope, i.e., the amount of change necessary for all students to be proficient or above by 2013-14 (the NCLB goal). The School Change score/grade is based on a school's performance against its own target slope, i.e., a slope line depicting the change in percent proficient from one year to the next. The method of calculating the individual target slopes for Michigan schools is set forth in Appendix A.

The committee recommends the following assignment of cut scores/grades for School Change

Elementary Reading Change				
Grade	Change Slope Range	Number of Schools	Percent of Schools	Cumulative Percent of Schools
A	125% of target	295	16.6%	16.6%
B	75% to 125% of target*	205	11.6%	28.2%
C	25% to 75% of target	330	18.6%	46.8%
D	Between 25% of target and 25% of target below zero	315	17.8%	64.6%
F	More than 25% of target below zero	627	35.4%	100.0%

The change slope range is expressed in terms of a comparison of a school's calculated slope to the target slope for the school as recommended by the Accreditation Advisory Committee.

* The three year average of the percent proficient is above 100 minus the number of years remaining to the goal of 100% proficiency (88% for 2002-03); schools that score at high levels in School Status will be given this grade, i.e., a B, unless their change score would qualify them for an A.

Elementary Mathematics Change				
Grade	Change Slope Range	Number of Schools	Percent of Schools	Cumulative Percent of Schools
A	125% of target	936	52.8%	52.8%
B	75% to 125% of target*	239	13.5%	66.2%
C	25% to 75% of target	151	8.5%	74.7%
D	Between 25% of target and 25% of target below zero	146	8.2%	83.0%
F	More than 25% of target below zero	302	17.0%	100.0%

The change slope range is expressed in terms of a comparison of a school's calculated slope to the target slope for the school as recommended by the Accreditation Advisory Committee.

The three year average of the percent proficient is above 100 minus the number of years remaining to the goal of 100% proficiency (88% for 2002-03); schools that score at high levels in School Status will be given this grade, i.e., a B, unless their change score would qualify them for an A.

Middle School Reading Change				
Grade	Change Slope Range	Number of Schools	Percent of Schools	Cumulative Percent of Schools
A	125% of target	87	12.9%	12.9%
B	75% to 125% of target*	121	17.9%	30.8%
C	25% to 75% of target	185	27.4%	58.1%
D	Between 25% of target and 25% of target below zero	174	25.7%	83.9%
F	More than 25% of target below zero	109	16.1%	100.0%

The change slope range is expressed in terms of a comparison of a school's calculated slope to the target slope for the school as recommended by the Accreditation Advisory Committee.

* The three year average of the percent proficient is above 100 minus the number of years remaining to the goal of 100% proficiency (88% for 2002-03); schools that score at high levels in School Status will be given this grade, i.e., a B, unless their change score would qualify them for an A.

Middle School Mathematics Change				
Grade	Change Slope Range	Number of Schools	Percent of Schools	Cumulative Percent of Schools
A	125% of target	321	47.6%	47.6%
B	75% to 125% of target*	108	16.0%	63.6%
C	25% to 75% of target	117	17.3%	80.9%
D	Between 25% of target and 25% of target below zero	67	9.9%	90.8%
F	More than 25% of target below zero	62	9.2%	100.0%

The change slope range is expressed in terms of a comparison of a school's calculated slope to the target slope for the school as recommended by the Accreditation Advisory Committee.

* The three year average of the percent proficient is above 100 minus the number of years remaining to the goal of 100% proficiency (88% for 2002-03); schools that score at high levels in School Status will be given this grade, i.e., a B, unless their change score would qualify them for an A.

Middle School Science Change				
Grade	Change Slope Range	Number of Schools	Percent of Schools	Cumulative Percent of Schools
A	125% of target	15	2.2%	2.2%
B	75% to 125% of target*	38	5.7%	7.9%
C	25% to 75% of target	165	24.6%	32.5%
D	Between 25% of target and 25% of target below zero	360	53.7%	86.1%
F	More than 25% of target below zero	93	13.9%	100.0%

The change slope range is expressed in terms of a comparison of a school's calculated slope to the target slope for the school as recommended by the Accreditation Advisory Committee.

* The three year average of the percent proficient is above 100 minus the number of years remaining to the goal of 100% proficiency (88% for 2002-03); schools that score at high levels in School Status will be given this grade, i.e., a B, unless their change score would qualify them for an A.

(On December 12, 2002, the State Board of Education adopted the above recommendations on cut-scores and grades.)

Grade 8 Social Studies

While five years of achievement data are desirable for the computation of the calculated slope, as an interim solution the measurement of social studies achievement change will be calculated on the basis of the four years of data that are available.

Grade 11 Reading

While five years of achievement data are desirable for the computation of the calculated slope, as an interim solution the measurement of high school reading for achievement change will be calculated on the basis of the three years of data that are available.

Grade 11 Writing

While five years of achievement data are needed for the computation of the calculated slope, as an interim solution the measurement of high school writing for achievement change will be calculated on the basis of the three years of data that are available.

Grade 11 Mathematics

While five years of achievement data are needed for the computation of the calculated slope, as an interim solution the measurement of high school mathematics for achievement change will be calculated on the basis of the three years of data that are available.

Grade 11 Science

While five years of achievement data are needed for the computation of the calculated slope, as an interim solution the measurement of high school science for achievement change will be calculated on the basis of the three years of data that are available.

Grade 11 Social Studies

While five years of achievement data are needed for the computation of the calculated slope, as an interim solution the measurement of high school social studies for achievement change will be calculated on the basis of the three years of data that are available.

For Student Growth

The valid and reliable measurement of Student Growth as called for under Education YES! including the attribution of growth, or the lack thereof, to individual schools present a number of problems. Not the least of these is the question of how to deal with the effects of high and uneven student mobility both within a given school and between a "feeder" and a "receiving" school. In addition, there are psychometric concerns. To validly and reliably measure growth, four requirement are necessary: (1) the domain is specified; (2) there are measures at adjacent grade levels; (3) these measures are equated with one another and reported on a vertically equated scale, and (4) a longitudinal tracking system based on verified demographic data has been created and a sufficient number of students remain in the cohort for comparison purposes.

Elementary and Middle School Levels

The Long-Term Solution

Since the requirement of measures at adjacent grade levels has not yet been met, and likely will not be met until the implementation of testing for *No Child Left Behind* begins, currently set for 2004-05 in Michigan, the committee strongly recommends that the State Board consider delaying the calculation of a student growth measure until such time as the state creates a cross-grade MEAP score scale as an integral part of the new assessment system being developed to comply with No Child Left Behind (NCLB). This will permit growth per year to be first reported in a valid and reliable manner following the 2005-06 test administration. The cross-grade scale should include equating Grade 11 MEAP measures with the respective Grade 7 and Grade 8 MEAP measures. It is our understanding that current plans do not call for linking the high school assessment to the grade 3-8 scale. This adds additional challenges to reporting growth from middle school to high school. These challenges are discussed at length in a following section on *High School Level.* But first, we turn to the *Elementary and Middle School Levels.*

If the recommendation to delay calculation of a Student Growth score at the Elementary and Middle School Levels is accepted, the Committee suggests that the weights for the two achievement measures, School Status and School Change, be set at 34 and 33 respectively.

The Committee's specific recommendation for the Elementary and Middle School Levels are spelled out in Appendix A.

An Interim Solution

If the State Board agrees with the longer term solution yet rejects the recommendation to delay and is committed to reporting Student Growth during the interim period, the Committee recommends the interim solution which also is spelled out in Appendix A.

Under the interim solution, the resulting single cut score/grade would be assigned to both the "feeder" and "receiving" schools (i.e., the 4th grade and 7th grade pair) since one cannot validly attribute student growth to either separately.

(At its meeting of December 12, 2002, the State Board of Education took action to implement the committee's proposed interim solution, rejecting the committee's recommendation to delay calculation of a Student Growth score at the Elementary and Middle School Levels.)

Since the State Board opted for the interim solution, the Committee recommends the following assignment of cut scores/grades for Student Growth. Following each recommended assignment the distributions of cut scores/grades for Student Growth for 2002-03 are also provided.

Elementary to Middle School Growth in Reading				
Grade	Growth Score Range	Number of Schools	Percent of Schools	Cumulative Percent of Schools
A	Above 33	266	15.4%	15.4%
B	29 – 32	611	35.4%	50.8%
C	25 – 28	442	25.6%	76.4%
D	20 – 24	273	15.8%	92.2%
F	Below 19	135	7.8%	100.0%

The growth score range is expressed in terms of the "interim solution" recommended by the Accreditation Advisory Committee.

Elementary to Middle School Growth in Mathematics				
Grade	Growth Score Range	Number of Schools	Percent of Schools	Cumulative Percent of Schools
A	Above 44	376	21.8%	21.8%
B	38 – 43	452	26.2%	48.0%
C	32 – 37	475	27.5%	75.5%
D	25 – 31	288	16.7%	92.2%
F	Below 24	135	7.8%	100.0%

The growth score range is expressed in terms of the "interim solution" recommended by the Accreditation Advisory Committee.

(On December 12, 2002, the State Board of Education adopted the Interim Solution and the above recommendations on cut-scores and grades.)

High School Level

As we noted above, since current plans do not call for linking the high school assessment to the grade 3-8 scales, there are a number of additional challenges that the State Board would face in calculating and reporting growth from middle school to high school. The situation at the high school level is much different than at the elementary and middle school levels. At the high school level, unless MEAP plans change, the current temporary situation at the elementary-middle school level (lack of test data at adjacent grades and the lack of a link between middle-

school and high-school test) will persist indefinitely at the high school level. We believe that the State Board of Education should carefully consider whether it wishes to report Student Growth scores from middle school to high school, since the best way to do so would be to build such an option into the MEAP program as it undergoes the current changes being made to comply with NCLB (i.e., adding grade-level benchmarks for mathematics and language arts/reading; developing tests for grades 3-8; and creating a cross-grade scale).

One way to build such an option into the MEAP would be to add every-pupil testing in grades 9 and 10, based on benchmarks in mathematics and language arts/reading for grades 9, 10, and 11 that are conceptually-linked to the benchmarks for these subjects at grades 3-8, and create a vertical scale from grades 7/8 to 9, 10, 11. This would permit the most direct comparison of performance of middle schools to high schools, or perhaps, the growth from grade 9 to grade 11 The obvious disadvantage of this approach is the added costs (time and money) for additional testing at two grade levels, neither of which are currently required or mandated. Therefore the committee does not see this as desirable or a feasible solution.

The committee also considered a number of other possible approaches to meeting the challenges that the State Board would face in calculating and reporting growth from middle school to high school. These possible approaches are spelled out in some detail in Appendix A. However, each of these approaches presents a number of serious problems, leading to grave doubts about their feasibility or desirability.

Recommendation. The committee has serious doubts about whether any of the solutions outlined in Appendix A are sufficiently sound or feasible. Before the State Board of Education even considers further any one of these solutions, it is our recommendation that the match issue first be resolved. If that match is sufficiently high for the state as a whole and for most schools, then a decision could be made on which method of linking might be feasible and desirable. Without such a match study to determine that the match is sound in a sufficient number of schools, the committee would recommend against reporting a Student Growth score for high schools and their feeder middle schools. In that instance, we believe that the data would be so flawed that it should not be reported.

In addition, if a MEAP solution is desired (solution one), we recommend that the State Board of Education give serious consideration to building in linkages *now*, where possible, within the MEAP system currently under redevelopment. Grade-level benchmarks, assessments, and cross-grade scales developed for the middle school and high school levels *as the assessments are created* is a much better solution than jury-rigged statistical procedures implemented after the fact.

Issues in Reporting School Status, School Change and Student Growth

The calculations of the cut-scores and grades presented for review and approval to the State Board of Education in December 2002 were based on results from 2001 and prior years.* In the case of the School Status cut-scores and grades, the calculations were based on the mean of three

*While the 2002 MEAP results became available to the Department in late summer 2002, they were not incorporated into the computation of cut-scores and letter grades in time to use in the standard-setting process.

years of results of *comparable* data from 1999, 2000 and 2001. In the case of the School Change cut-scores and grades, the calculations were based on the means of five years of *comparable* data from 1997, 1998, 1999, 2000 and 2001. The sole exception to this pattern was the 4th and 7th grade MEAP Reading assessments where results were available from 2002 at the time the cut-scores and grades were calculated.

The 2002 results, in all cases, are now available, and very soon the 2003 results will be available On the face of it, this appears to present the opportunity to update the calculations by including more recent results, i.e., both the 2002 and 2003 MEAP results. However, a serious problem now presents itself. With the exception of the Social Studies assessment, the 2002 and 2003 results are not directly comparable with prior years results, making the calculation of 3-year MEAP score averages for School Status and the calculation of three 3-year MEAP score averages for School Change problematic, to say the least.

In the case of Reading, new MEAP assessments at grades 4 and 7, unequated with the Reading assessments given in prior years, were administered in the 2002-03 school year. In the case of Mathematics and Science, new MEAP assessments at grades 4 and 7, again unequated with the assessments given in prior years, were administered beginning in the 2001-02 school year. In short, the results from 2002 and 2003 cannot be joined with the results from prior years to calculate the needed 3-year MEAP score averages for School Status and the three 3-year MEAP score averages for School Change.

Until such time as three and five years of comparable data are available, or until a successful equating process is undertaken and completed, it will be virtually impossible to calculate and report these cut-scores and grades using the most recent results. Furthermore, since the 2002 and 2003 MEAP Reading results are not comparable, the computation of a Student Growth score from elementary to middle school also becomes problematic.

However, there may be an interim solution that will allow the reporting of cut-scores and grades, albeit in a less than totally satisfactory manner.

An Interim Solution

Under this interim solution, School Status for Reading would utilize only comparable data from 2000, 2001 and 2002 until such time as at least two years of comparable data from the new Reading assessments would be available, i.e., 2003 and 2004 results (picking up the full three years of comparable results in 2005). School Status for Mathematics and Science would be limited to 2-year MEAP averages until the addition of the 2004 results. While not an ideal solution, since the 2-year MEAP score averages will be less stable than the 3-year averages, it is a workable solution.

The situation is more problematic for the School Change cut-scores and grades that ideally should be calculated using three 3-year averages, or five years of comparable data. Under the interim solution, School Change for Reading would utilize only comparable data from 1997, 1998, 1999, 2000, and 2001 until such time as at least three years of comparable data would be available, i.e., 2003, 2004 and 2005 results (picking up four years and then the full five years of

data in 2006 and 2007 respectively). School Change for Mathematics and Science would be limited to 5-year MEAP averages from 2001 and prior years until the addition of the 2004 results, which would provide 3-year MEAP averages (moving to 4-year and 5-year averages in 2005 and 2006). Again, while not an ideal solution, since the 3-year MEAP score averages will be less stable than the 4- and 5-year averages, it is a workable solution.*

The Student Growth measures require comparable results from two consecutive years. Consequently, reporting Student Growth for Reading in 2003 would be limited to utilizing comparable results from 2001 and 2002. In 2004, since two years of comparable results will be available, the problem will be eliminated and the reports would be based on 2003 and 2004 results. Comparable results for Mathematics and Science are available for 2002 and 2003, hence there is not a problem.

Below is a graphic display of the proposed interim solution

MEAP results used in calculating and reporting Education YES! cut-scores & grades

Subject Area	In 2003	In 2004	In 2005	In 2006	In 2007
Reading					
Status	00-01-02	03-04	03-04-05	04-05-06	05-06-07
Change	98-99-00-01-02	98-99-00-01-02	03-04-05	03-04-05-06	03-04-05-06-07
Growth	01-02	03-04	04-05	05-06	06-07
Math & Science					
Status	02-03	02-03-04	03-04-05	04-05-06	05-06-07
Change	98-99-00-01-02	02-03-04	02-03-04-05	02-03-04-05-06	03-04-05-06-07

The proposed interim solution as well as the cut-scores and grades calculated under the solution will need to be reviewed and approved by the State Board of Education. It also will be essential to inform the schools of the bases on which these scores will be calculated, i.e, the years from which the MEAP results are drawn and the rationale for doing so.

The Use of a Standard Setting Panel in Setting Cut-Scores and Grades

Standard-setting is the term used to label the judgmental process that leads to establishment of cut scores or grades in certain assessments. Standard-setting is always, in every field of endeavor, a judgmental process. There is no one correct or "safe" way to do it.

For future years, as new MEAP assessments are developed and come onto line, it is imperative that the MEAP office take the steps necessary to equate newly developed assessments with prior year assessments to ensure comparability of results over time. Unless this is done, the problem described above will continue to plague the calculation and reporting of Education YES! results in future years. While the proposed interim solution offers a means of addressing the immediate problem, i.e., Education YES! reporting in 2003 and 2004 , the problem also could be addressed if the Department, or the MEAP Office, were able to undertake and successfully complete an immediate equating of the 2003 and 2002 results with the results of prior years.

The process of developing performance standards usually includes three phases: (1) defining the standards with descriptive language that communicates the differences among the performance levels; (2) setting *weights* for the individual components (or scores/grades) that make up the aggregate performance standard; and (3) making *judgments* about the cut-scores/grades to be assigned to different levels of performance in the individual components, and for aggregate performance.

In *Education YES!,* the second phase was done by policy fiat, i.e., the State Board set the weights to be assigned to the individual components that would make up a school's aggregate score/grade. On a total weighting of 100 points, 33 points were assigned to the eleven indicators of performance that aggregate into three major areas (Engagement, Instructional Quality, and Learning Opportunities); the remaining 67 points were assigned to the three achievement measures: School Status, School Change, and Student Growth. The responsibility for the third phase was assigned to the Accreditation Advisory Committee.

To fulfill its charge, the committee chose to recommend a formal standard-setting process for several reasons. First, the wide variety of variables on which schools are to be graded (achievement and school performance variables) suggests that weighting among these variables would need to be considered. Second, the grading of schools is controversial. Third, there is relatively little experience in setting overall grades for schools based on such variables. Finally, it is the committee's desire that attention shift from the process of determining the grades to helping schools improve their performance. One way to do this is to try to set standards for schools through the most inclusive process possible.

Careful conceptualization and implementation of the process requires that the judges examine actual score/grade profiles. The role of the judges is <u>not</u> to set the standard but rather to recommend cut scores/grades to the policy makers, whose responsibility it is to decide the actual cut scores/grades, considering the recommendations from the judges as well as other factors that may legitimately influence their final decisions.

Consequently, the committee asked Department staff to convene a standard-setting panel broadly representative of teachers, administrators, parents, and members of the business community. The panel met on two separate occasions. In a two-day work session in late September 2002 and following a carefully guided procedure, the panel examined actual score profiles for 4th grade reading and mathematics, and middle school reading and mathematics, and based on that examination offered its recommendations on cut scores/grades for School Status and Student Growth to the Accreditation Advisory Committee for its review. In a second meeting, in mid-October 2002, the panel met again to examine score profiles for middle school social studies and science and offered its recommendations on cut scores/grades for School Status and Student Growth to the Committee. In the view of the Accreditation Advisory Committee, this process was a sound one, one that adhered to the standards of the measurement community, and one that will ensure the soundness, validity acceptability of the cut-scores and grades to be assigned to Michigan schools.

The accreditation committee, based on its review of the work of the standard-setting panel (and on its own judgments with respect to the School Change cut scores/grades), recommends the set

of School Status, School Change, and Student Growth scores/grades identified above to the State Board of Education. The State Board, of course, has final responsibility to set the actual cut scores/grades to be assigned to schools*.

A complete report of the standard-setting process is available from the Department.

II. Measuring School Performance Indicators

The Initial Development of the Measures

The Accreditation Advisory Committee fully supports the State Board's goal of including performance indicator data in addition to MEAP data in Education YES! Performance indicator information coupled with information on academic achievement will present a full and complete picture of the performance of Michigan's public schools. However, the development and incorporation into Education YES! of a valid and reliable set of performance indicator measures from which cut-scores and letter grades can be drawn is a daunting task, and a task which is not easily accomplished in the short period of time that was allotted to it.

These realities led the committee to initially recommend that the planned January 2003 launch of performance indicator data collection efforts be seen, at best, as a first step in a needed three-year developmental process. This developmental process would have as its goal the generation, refinement and eventual adoption of a set of valid and reliable program indicator measures. Once such a set of measures was developed, these measures and their accompanying cut-scores and grades could be established with confidence as an integral part of *Education YES!*

We further noted that the development of the current performance indicator measures proceeded on a separate track relatively immune from feedback offered by the Accreditation Advisory Committee. The performance indicators were developed by teams of intermediate and local district educators largely in isolation from the work of the committee.

The 30 School Pilot

In late October 2002, the indicator measures were tried out in 30 schools spread out across the state, each of which completed the measures in writing and then were visited by a team to verify the ratings in person.

Using grades to report school performance, the State Board may decide whether to use typical school definitions of the grades to be assigned. For example, an "A" means that a school did outstanding, while a "B" designates a school that is above average, a "C" a school that is average, a "D" a below-average school, and an "F" a failing school. Standard setters might embellish these definitions with other aspects of these schools, depending, for example, on a school's performance on the school performance indicators.

In early December 2002, the Accreditation Advisory Committee examined the data collected in the 30-school pilot study. This examination suggested to the committee that there well might be serious problems with the performance indicator measures as presently crafted. First, if the data collected in the full-scale implementation planned for January 2003 were to have the same characteristics as the pilot data, the results would not accurately reflect the variation in quality among Michigan schools. Second, not only would there be little variation among schools but the great majority of schools likely would score very high on the measures, leading to a distribution skewed largely to the right. Third, based on the results of the pilot study, there was little reason to believe that the data from the full-scale implementation would produce valid and reliable measures on which to assign cut scores and grades; most schools likely would end up with high scores on each of the three major indicator areas, thus making it virtually impossible to assign grade levels (A,B,C,D,F) based on the indicators. If all schools end up with similar high scores, how would this help all schools to lift themselves up to improve school achievement? Fourth, rather than measures of performance, the performance indicators, as crafted, fell more appropriately into the category of measures of program policies. Very few, if any, of the indicators were oriented toward outcomes. We believe they should be.

Still, the committee believed that the results of the full-scale implementation in January 2003 might help guide needed modifications for Year 2. But the committee also expressed the view that more than that is needed. A different approach, based on outcome measures of each indicator, set against clear performance benchmarks, will need to be developed. One possible starting point for this redevelopment effort is the initial work done by the group at Michigan State University in Spring 2002 to provide a research basis for each indicator. A small group of advisers composed of persons who have successfully undertaken similar efforts could contribute needed guidance and direction to such an effort. This will lead to the redevelopment of a set of valid and reliable measures that could serve as the basis for setting cut-scores and grades in this vital area of Education YES! in the years to come.

The committee also felt that preliminary information on results from the January 2003 data collection effort still could be reported *if warranted*. However, the committee recommended that until such time as an effort akin to that described above was undertaken and resulted in the production of valid and reliable measures, no performance indicator scores and grades should enter into the calculation of the composite school scores and grades.

The Results of the Full-Scale Survey

In early February 2003, the Accreditation Advisory Committee reviewed the School Performance Indicator results following the completion of the January 2003 school self-assessment process (which involved 3,844 Michigan schools). The committee's worst fears were realized. In examining the schools' ratings, approximately 60 percent of the schools gave themselves ratings of 90 percent or higher of the possible points on each cluster of indicators. The committee is concerned that there is insufficient variation in the results to reliably grade schools on these indicators. Hence the committee's initial recommendation, as noted above,

would be not to include the School Performance Indicators in Education YES! until such time as improved measures are available.

An Interim Solution

However, we also are mindful that such indicators should and could be a part of a system to accredit schools. Therefore, we propose for the State Board's consideration the following interim solution:

A. Calculate the percentage of possible points each school gave itself on each of the 11 school performance indicators.
B. Calculate the error of measurement for each indicator.
C. Schools that scored above 100 % minus 2 standard errors of measurement (2 standard errors of measurement below 100%) on an indicator would receive 1 point for that indicator.
D. Schools that scored below that level (100 percent minus 2 standard errors) on an indicator would receive a score of "0" on that indicator.
E. Report scores on each indicator (0 or 1) as well as the sub-total scores on the 3 clusters of indicators, and the total score. No grades would be reported for any of these scores.
F. School Performance Indicators would count for only 11 points (not the 33 points as approved by the State Board) until such time as the improved measures are available.
G. Keep MEAP achievement score weightings the same (School Status @ 23 points; School Change @ 22 points; Student Growth @ 22 points). Add 11 points for School Performance Indicators to produce an interim total weighting of 78 points.
H. Once School Performance Indicator measures are improved, raise the contribution of these measures back to the original 33 points.

The advantages of this interim solution are:

1) It recognizes the work of districts in completing the Indicator measures.
2) It keeps the measures in Education YES! (albeit with a lesser weighting).
3) It encourages continuing work to improve the measures, but does not place undue weight on the current measures.

III. Calculating the School's Composite Score

As set forth in Education YES!, a composite school score set on a scale from 0 to 100, along with a corresponding letter grade of A, B, C, D or F, is to be reported for each public school in Michigan. Schools that receive an A, B, C or D-Alert will be accredited. Schools that receive an A will be summary accredited. Schools that receive B, C or D-Alert will be in interim status. Schools that receive an F will be unaccredited.

The General Procedure

Education YES!, through its pre-set weighting scheme, sets forth the method of calculating the composite score for any given school. Under that weighting scheme, from a total weighting of 100 points, 33 points are assigned to the eleven indicators of performance that aggregate into three major areas: *Engagement, Instructional Quality*, and *Learning Opportunities*; the remaining 67 points are assigned to the three achievement measures: *School Status, School Change*, and *Student Growth*. The six major components to which grades are to be assigned and their weights are:

Component	Weight
Achievement	
School Status	23
School Change	22
Student Growth	22
Performance Indicators	
Engagement	9
Instructional Quality	12
Learning Opportunities	12
Total	100

In calculating a school's composite score, first the subject area scaled scores for each of the achievement measures for the school are transformed to scores on a 0 to 100 scale, then averaged to produce a single score; for example, the index scores for *4th grade Reading* and *4th grade Mathematics* are transformed to scores on a 0 to 100 scale, then averaged to produce a single score for *School Status*. The process is repeated for *School Change* and *Student Growth*. Then the scaled scores for the sub-components of each of the three performance indicator areas are transformed to scores on a 0 to 100 scale, then averaged to produce a single score for each area; for example, the scaled scores of the sub-components *performance management systems*, *continuous improvement*, and *curriculum alignment* are transformed to a 0 to 100 scale, then averaged to produce a single score for *Engagement*. The process is repeated for *Instructional Quality* and *Learning Opportunities*. The resulting scores for the six major components listed above are then multiplied by their respective weights, totaled, and divided by 100 to give, on a 0 to 100 scale, a single score for the school. This single score is then transformed to the appropriate corresponding letter grade, e.g., A or B or C or D or F.

A complete description of the actual method for calculating the composite school score, including the transformation from the scaled score to a score on the 0 to 100 scale, and then transforming that score to a letter grade is set forth in Appendix B.

The Problem of "Missing Data"

The Department will need to decide how to handle accreditation decisions for schools that, for legitimate reasons, have "missing data" and for which no composite score, at least as described above, can be calculated. These schools will fall into one of two categories: (1) the school that consists only of grades at which no MEAP scores currently are available, e.g., a K-3 school; and (2) a school that is newly formed and as a consequence does not have the minimum of 3 years of MEAP scores required to calculate Status, Change and Growth scores, e.g., a newly formed charter school.

IV. Alignment of Education Yes! with Federal Legislation

Setting the Baseline for AYP

At its June meeting, the Committee met with Department staff and, following a review and discussion, supported the staff's initial recommendations for the 2002-03 school year, namely

Use total scores on only Reading and Math for triggering sanctions

- Use NCLB target of the 20th percentile for the baseline
- Report all results under the AYP format
- Urge schools to disaggregate and review/analyze their data in anticipation of using sub-populations as sanction triggers for 2002-03

A baseline set at the 20th percentile school (20th percentile of the State's total student population) seemed appropriate and reasonable to the commitee—on average about a 40% proficiency level, i.e., some 60% are not proficient.

Templates of school results could be overlaid on a graph of the state AYP (irrespective of where the state baseline is eventually set), thus relating or referencing a school's Change score/grade to the AYP "score."

(Subsequently, the State Board of Education took action to adopt the staff recommendation with one change, namely, to require schools to use sub-populations as sanction triggers in 2002-03.)

Tying Education Yes! to NCLB

The committee also considered an approach advanced by Department staff to tie the two systems even closer together. Essentially the approach would use the School Status score, and its assigned grades, to set the bar for proficiency e.g., a grade of D or F on the *Education Yes!* School Status score would trigger the AYP sanction unless the School Change grade was B or better (the federal "safe harbor"). The bar would be raised from year to year, or from a block of years to a block of years, so that a D, then a C, then a B, and on to an A, in effect, would become the bar -- gearing all of this to the expectation of 100% proficiency at the end of 12 years.

The committee is in general agreement with linking *Education Yes!* and AYP. It is not of a single mind about how tight the link ought to be. On the one hand, two members argue that, while closely related, *Education YES!* and NCLB are separate systems serving similar but somewhat different purposes. To too tightly link *Education Yes!* to NCLB would mean that the success of *Education Yes!* would become very closely intertwined with NCLB's success (and lasting power?). On the other hand, two members argue that this tight link is in accord with a major purpose of *Education YES!,* namely ensuring that all students reach proficiency—and in a set period of time, i.e., the 12 years of NCLB.

At root, how tight the link ought to be is a policy decision and one that the SBE is best able to consider and make. Department staff certainly can, and will present, the particulars for the State Board of Education.

(On November 14, 2002, the State Board of Education took action to link Education YES! and NCLB, and to amend the Education YES! document to reflect the nature of that linkage. The State Board action provides that a school that does not make AYP shall not be given a grade of "A." A school that makes AYP shall not be listed as unaccredited. A school's composite grade will be used to prioritize assistance to underperfoming schools and to prioritize interventions to improve student achievement. (See p. 6 of *Education YES! – A Yardstick for Excellent Schools*, November 14, 2002.)

Focus on Performance in Mathematics and Reading

The Accreditation Advisory Committee recommends that *Education YES!* grades for accreditation include Mathematics, Reading, Science and Social Studies. However, the Committee recommends that for the calculation of AYP under NCLB, that AYP should focus on performance of students in the areas of mathematics and reading. We recommend this change for the following reasons:

1. NCLB Title I focuses primarily on the achievement of students in mathematics and reading;
2. NCLB is intended to give parents/guardians of students most at risk because of low mathematics and reading scores the opportunity for their children to receive additional instructional support or to move their children to a school not in need of improvement in these areas;
3. Adding schools to the "in need of improvement" category because of low performance in science (or social studies) will greatly increase the number of schools deemed to be "in need of improvement;"
4. Increasing the number of schools "in need of improvement" will greatly reduce parental options to seek help for children in mathematics or reading;
5. While mathematics and reading are not more important than science or social studies, in order to meet the NCLB goal of 100% proficiency in mathematics and reading in twelve years, schools, districts, and the state must be very pointed and focused with their efforts and resources;
6. The addition of science and social studies to NCLB AYP will dilute the resources and efforts needed to help schools achieve universal proficiency, thereby increasing the

likelihood of failing to help all students achieve at high levels in mathematics and reading; and

7 Projections and calculations of schools "in need of improvement" in Michigan have not yet addressed subgroup performance, which are anticipated to greatly expand the number of identified schools.

(The November 14, 2002, State Board of Education action to link Education YES! and NCLB incorporated this recommendation.)

V. The Initial Year of Statewide Implementation and Further Development of the System

The committee recommends that 2002-03 be viewed as "the first year of statewide implementation" of Education YES!, subject to further study and refinement over the course of the next two to three years, especially in terms of identifying measures, indicators, and additional areas that will need particular attention.

We strongly recommend that the following actions be taken

1 First, we believe that one of the keys to both the initial and long-term success of Education YES! is the manner in which the results are reported to educators and other members of the public. We know of successful communication efforts planned and undertaken for other large, innovative educational programs (e.g., The National Assessment Governing Board's efforts to better communicate NAEP results, and the 1989 Michigan Department of Education campaign to explain the new MEAP Reading test and anticipated lower test performance). We strongly urge the Department to take steps to learn from these efforts and to apply the lessons learned to Education YES! as it is rolled out in its first year. Furthermore, we urge the Department to build on the lessons and the experiences of the first-year rollout in improving the manner of reporting in year two and subsequent years.

2 Second, there are a number of immediate actions that should be taken by the Department during the first year, with a view to identifying modifications that should be made prior to the rollout of Education YES! in its second year (as well as in subsequent years). These include the following:

- Since time did not permit a comprehensive review of Education YES! to be conducted prior to first year reporting, we suggest a more complete review process be carried out prior to the second year rollout. This review, at a minimum, should include:

  - cross-validation by a third party of the first year results

    studies using the assigned grades to investigate the quality of the scales used for the school performance indicators. Since some of

these are new and, in some cases, provide measures of school attributes not previously measured, the scales used to collect the data need to be further investigated.

studies on the manner in which the school performance indicators are measured and combined to yield school grades, since the set of variables used is unique and has not been used for such purposes in the past.

with respect to the Student Growth score, appropriate technical adjustments should be explored and implemented to address the problem of high growth at the lower grades and less growth, or plateauing of growth, at the higher grades.

- It is our belief that the more care that is taken in how the various school performance measures and the student achievement measures are combined into School Status, School Change, Student Growth and Performance Indicator scores and grades, and ultimately into a Composite School score and grade, the greater will be the support provided by educators and the public for the grades that are assigned. Correspondingly, this also will lead to less time being spent on squabbles over the grades and more time on improving building-level performance. Two particular areas that need further investigation are (1) the conversion of scores and grades to the 0-100 scale, and (2) the calculation of the composite school score and grade—both of which are called for in Education YES!

  A small set of schools should be randomly-selected to be visited in person to review their school performance indicator information and MEAP scores to determine: (1) how accurately they described the school; and (2) whether the scores and grades that the school received are accurate in the eyes of the community, school board, administrators, teachers, and parents.

  A procedure should be established for identifying schools that have developed successful approaches to communicating an understanding of Education YES! to the school community, and have used data and information from Education YES! to initiate actions aimed at school improvement. These visits should include an exploration of what technical support might be helpful to the school in its continuing efforts to raise proficiency levels.

  An appeals process should be established so that any school or person who receives the school's grades, or receives reports of them, may appeal the results. Appeals should be based on a clear and sound rationale. For example, an appeal could be based on a contention that the grade(s)

assigned are too high or too low, given what the appellant knows of the school situation. Individual instances should be reviewed, and areas of potential inaccuracy tracked to determine whether there are broader issues that need further review and possible modification.

Once these steps have been taken, recommendations for appropriate modifications in the accreditation system should be presented to the State Board of Education for approval and implementation in the 2003-04 school year. Periodically following the year 2003-04 (e.g., on an annual basis), we recommend the State Board of Education re-visit the system to make sure that it continues to work well and that further modifications, as needed, are incorporated.

4 To this end, we recommend that the State Board of Education and the Department consider establishing a small standing panel of respected experts in measurement, evaluation and statistical analysis to provide the State Board and the Department with technical advice, as well as policy advice, on further development and implementation of Education YES! The panel's general charge would include periodic and ongoing reviews of the validity and reliability of the key measures used in Education YES! and the improvement of these measures over time. In addition, the panel would evaluate and advise on the usefulness of the system for reporting publicly on school performance, as well as the system's utility in advancing ongoing reform efforts. It is further recommended that the members of this panel be drawn from other than current members of the Accreditation Advisory Committee. The panel might meet as a group no more that two to three times a year; a good portion of the panel's work could be accomplished without necessarily bringing the members together.

5. There is a pressing need for Education YES! and MEAP to be "on the same page," and overseen by a single policy executive and policy body. These two programs should be tightly aligned. The current locations of Education YES! and MEAP in two separate principal departments of state government, Education and Treasury, not only make little organizational sense, but also give rise to acute problems of coordination between the two programs. A most recent example, described in detail in pages 12-14, is the problem that has arisen in the reporting of Education YES! as a consequence of the non-comparability of MEAP results over multiple year spans.

6 There is a need for technical documentation on the validity of MEAP assessments for use for accreditation purposes under Education YES! The Education YES! system includes a number of components that rely on MEAP data for determining the school grade including the Status, Change, and Growth measures. For these components of the system to work well, the MEAP tests need to have technical characteristics that support the inferences that will be made from the test results. These technical characteristics mainly deal with the need to have tests that can accurately reflect the change in school mean scores and individual student performance. All tests are not equally good supporting particular judgments from the test scores. An easy test will not be very useful for selecting the top 1 percentof candidates for scholarships and a

hard test will not be useful for diagnosing learning difficulties. In this case, for Education YES! to function properly, the tests must be sensitive to changes along the range of performance that is the focus of the accreditation program. This range is from the current NCLB cut-score to the definition of proficient. This is a fairly wide range and it is important that evidence be provided to show that the tests can accurately represent change along that full range. If the tests are insensitive to change, it will appear that schools are not making progress toward their goals when in fact they are. If that is the case, the features of the tests can undermine the goals of the accreditation program.

7 Finally, the committee strongly urges the State Board and the Department, even in this time of serious revenue constraints, to adequately resource and staff Education YES! This might best be done by setting as a top priority the establishment of an accountability unit within the Department, providing that unit with adequate human and financial resources, and charging it with the further development and implementation of Education YES! This unit also should have the wherewithal to engage a statistical contractor to assist it in carrying out its responsibilities, including the preparation of detailed technical documentation on the development and implementation of the program. Adequate resources are essential to the success of the program. If a program of the complexity of Education YES! is carried out on a "shoestring," it likely will collapse of its own weight.

NEXT PAGE