Research Report 1
Feb. 2002

Building Valued-Added Assessment into Michigan's Accountability System: Lessons from Other States[1]

Kwangyhuyn Lee, Policy Analyst, The Education Policy Center at MSU
Debbi Weimer, Policy Analyst, The Education Policy Center at MSU

Introduction: The MEAP is not Enough

Michigan is designing a new accountability system that combines high standards and statewide testing within a school accreditation framework. It is not an easy process. Since a number of states have already developed accountability systems, Michigan can use their experiences to evaluate what works and what doesn't. Mathers (1999, 1) provides a standard by which Michigan can judge the accountability systems used by other states.

> An accountability system is complete when teachers, students, building and district leaders have clear instructional goals (*standards*), when states and local districts have developed *sound assessment techniques* and *quality indicators*, and when visible consequences for all involved parties have been put into practices (*rewards and sanctions*).

Sound assessment techniques are critical if the accountability system is to provide relevant information to schools and policymakers. One important component of a sound assessment system is measurement of student learning during the school year. There are two primary ways to assess student learning, absolute measures of achievement and value-added assessment. For reasons discussed in more detail in Section I, value-added assessment generally provides teachers, parents, educational leaders, and policymakers with a clearer picture of school effectiveness than absolute measures of achievement.

Value-added assessment attempts to distinguish learning that occurs because of factors that the school can control (such as teacher quality, academic rigor, and alignment with standards) from learning that is affected by uncontrollable factors (such as student poverty and community support). When we measure the value that was added to a student's achievement by a school, rather than by factors outside of the school's control, a school can be held accountable for the performance of students in that building. Value-added assessment was pioneered on a statewide basis by Tennessee and by the Dallas, Texas school district. Since that time, many other states and districts have begun moving towards a value-added system.

To measure the value added by schools and teachers, there must be annual testing in every grade and we must be able to equate these scores to the same scale each year. The changes in a student's test scores over time then form the basis for assessment of progress. Value-added testing allows factors that are not under the teacher's or school's control to be filtered out through longitudinal analysis of the test results[1]. With value-added testing, a student's socioeconomic status (SES) can be filtered out by looking at how much each student's score changes over time. Since a student's socioeconomic status and other family characteristics do not change dramatically from one year to the next, it is reasonable to think that each student's achievement will grow at an annual rate based in part on their SES and ability

[1] The specific form of the analysis varies. North Carolina, Tennessee, and Texas use a longitudinal method to analyze longitudinal data. Dallas uses longitudinal data to obtain residuals and then regresses the previous year's residual on the current year residual. All of the systems have longitudinal aspects but the longitudinal component is handled differently by the models used under each system.

In contrast, absolute measures of student achievement such as average scores provide a snapshot of learning but do not allow us to separate school and non-school influences on student achievement. MEAP tests, for example, provide us with a high quality tool for assessing students' knowledge. If we then try to evaluate schools based on their average MEAP scores, however, we will tend to reward schools that serve students whose home environments are more congruent with the school environment and punish schools whose students face more challenging circumstances. A snapshot analysis of MEAP scores cannot be used to compare schools that serve different populations, since research shows a close correlation between family income and MEAP scores.

In order to evaluate schools fairly, we need to incorporate value-added measures of achievement into our accountability program. Imagine that we have two schools that serve very different populations of students. School A might have very large gains but, because their students start out at a lower level of achievement, their absolute score may only be mediocre. School B may not add very much value to their students' learning but, because School B's students start out at very high levels of achievement, their absolute score may be as high or higher than school A's absolute score. Under a value-added assessment program, School A's excellent teaching and learning will be recognized. In contrast, School A's hard work might go unrecognized under a system that only uses absolute measures of achievement. It could even be labeled a failing school despite the outstanding job being done by School A's educators, administrators, students, and parents.

Michigan has already created the Michigan Education Information System (MEIS), an

infrastructure that gathers school data via the Internet, stores the information in a secure warehouse, and will soon make selected information available to appropriate decision makers. One key element of MEIS that is still under development is the Single Record Student Database. This database will allow information on individual students to be linked to teacher, fiscal, and performance data. For the first time, Michigan will have the infrastructure to track student progress from one year to another.

This paper is divided into three sections: Section I examines the background of value-added assessment and discusses why it should be part of Michigan's accountability system. Section II compares the implementation of value-added testing in several states and examines similarities and differences between the models. Section III outlines steps that would be necessary to implement a value-added system in Michigan.

Section I: Why Value-Added Assessment is Important

Initially used in evaluations of university programs (Taylor, 1985), the value-added concept was introduced into K-12 assessment systems in the 1990's. Cooley (1991) argued that districts needed to be held accountable for *improving* student performance, not performance *levels*. North Carolina, Tennessee, and the Dallas Independent School District (DISD) pioneered the introduction of value-added assessment. A brief overview of the literature highlights the advantages of value-added assessment:

- School incentive systems should be based on gains or improvement in student learning and the contribution of teachers and schools to those gains (Bryk, Deabster, Easton, Luppescu, & Thum, 1994; Hanushek, 1994; Hanushek & Meyer, 1996; King & Mathers, 1997; Ladd, 1999; Meyer, 1997). Hanushek (1994) argued that "schools and teachers should be held responsible only for factors under their control and rewarded for what they contribute to the educational process, that is, the 'value' they add to student performance." As of 1999, 19 states reward schools with money, flexibility, or recognition (*Quality counts 99*, 1999) based on some measure of student achievement. Value-added assessment systems are more likely to reward schools for their contribution to student achievement, rather than for factors outside of their control.

- To be fair, accountability systems must include performance measurements based on student gain and must consider student socioeconomic status (King & Mathers, 1997). King and Mathers (1997) further assert that the data analysis used to measure the student learning gains should be longitudinal. Considerations of fairness have encouraged some states to adopt or create value-added assessment systems.

- The most common aggregate indicators of performance, average and median test scores, are not valid measures of school or classroom performance (Hanushek & Meyer, 1996; Ladd & Clotfelter, 1996; Meyer, 1997). Comparing average test scores from year to year for a school will be misleading in urban

contexts with high annual student mobility rates and other factors[2] (Bryk et al., 1994).

- An additional problem is the volatility of test scores. Using single year measures of achievement introduces two main problems, sampling variation and one-time factors (Kane & Staiger, 2001). Sampling variation stems from year to year differences in the makeup of the group of students who are being tested. This is a serious problem in urban and other schools where student mobility rates are high. It is particularly severe when only a small number of students are tested each year. One-time factors might include problems such as an outbreak of flu during the week that tests are administered, or an uncomfortable temperature in the testing room. The "noise" created by these types of problems is often larger than the small changes in genuine school performance that are observed from year to year (Kane & Staiger, 2001). While a value-added system does not remove the influence of one-time factors, it does minimize sampling variation since each child serves as his or her own control. Pooling value added results across several years can help minimize the one-time factor problem.

The development of sophisticated statistical methods has facilitated the design of effective value-added assessment systems. Many states have sought to measure teacher effectiveness or school effectiveness in the last three decades, but most models relied on cross-sectional analysis using snapshot achievement scores as the criterion variables (Webster & Olson, 1988). There were few scholars who used student achievement gains to measure teacher and school effectiveness. During the 1980s, statisticians developed advanced techniques such as multilevel or hierarchical linear models. Measurement of school effectiveness using student test scores became a more reasonable approach thanks to these developments (Raudenbush & Willms, 1991). The Dallas Independent School District (DISD) and the state of Tennessee were two of the first systems to adopt sophisticated statistical models.

Section II: How Have States Instituted Value-Added Assessment?

Value added assessment systems vary dramatically. To illustrate the different systems, we will look at four accountability models that include a value-added component. Tennessee and the DISD use complex statistical models. Texas and North Carolina have developed relatively simple models that combine value-added measurement with measurement of absolute levels of student achievement.

It is also important to look at what tests are used to evaluate progress. Some states are interested in measuring how much their schools improve test scores relative to the gains on the national or state norm every year. These states' value-added assessment systems are *norm based* -- using either the national norm or the state norm. Other states are interested in measuring gains based on a criterion score set by the state for the year. These states' value-added assessment systems are *criterion based*.

[2] Adcock, 1995; and Wan, Haertel, and Walberg, 1993 also conclude that the simple average of student scores is not a good measure of school effectiveness.

The chart below provides a quick overview of the value-added assessment systems used in the Dallas Independent School District, Tennessee, Texas, and North Carolina.

System	Value-added vs. snapshot	Type of assessment instrument
Dallas	Both – absolute achievement and value-added measures	Both – Uses SAT9 (norm) and TAAS (criterion) tests
Tennessee	Both- value-added is emphasized (elementary schools must meet the 50 percentile national norm, gain)	TCAP – Customized CTBS test (norm tests)
Texas	Both – absolute is emphasized but various categories of students are expected to make comparable gains each year	TAAS – criterion-based tests
North Carolina	Both – value-added measures and absolute achievement. Emphasis on value-added	EOC and EOG – criterion-based tests

The Dallas Independent School District

The Dallas Independent School District (DISD) was the first district to incorporate a value-added model into its accountability system. The Dallas model was developed and continues to be administered under the long-term leadership of William Webster, the director of Research and Development at the Dallas Independent School District. In 1984, the DISD began using a school ranking system using multiple regression to develop longitudinal student growth curves. However, this model was abandoned when a new state accountability system was mandated. The current Dallas accountability system uses a combination of multiple regression and hierarchical linear modeling.

Dallas administers the nationally normed Stanford 9 as well as the TAAS (Texas Assessment of Academic Skills) to all students in grades 3 to 8. Dallas has a three-tier accountability system developed by William Webster (Webster & Mendro, 1995) beginning at the school level. The three tiers are:

1) School Improvement Plan (SIP)
2) District Improvement Plan (DIP)
3) School Improvement Indices

The first tier is the School Improvement Plan (SIP). The district provides each school with data; each school then establishes a five-year school improvement plan. The SIP targets include: student performance in language arts, math, social studies, and science; parental and community involvement in the schools; student promotion and course passing rates; student enrollment in advanced courses; diploma plans and honors plan; student graduation rates (dropout rates); student college entrance test participation and performance; student attendance; teacher attendance; and school climate and safety. Each school develops a strategic plan for each of the targets.

The second tier of the accountability system is the District Improvement Plan (DIP). The District Improvement Plan is built on the goals of the School Improvement Plans. The DIP determines absolute accountability objectives and specifies how central office divisions will support the individual schools. Both the SIP and DIP have absolute goals. This is sometimes problematic since the goals are subject to low expectations and are set without considering the reasonableness of goal attainment.

The last tier, School Improvement Indices, uses a methodology referred to as Dallas value-added assessment. The School Improvement Indices combine a regression model and hierarchical linear model (HLM) using longitudinal student achievement data. The Dallas method includes two stages: the first stage uses multiple regression analysis to determine how much of student achievement cannot be accounted for by student characteristics such as race and free lunch eligibility. What is left over is termed "residuals." In the second stage, a two-level hierarchical model is constructed by using the residuals obtained in the first stage. This stage of the model accounts for the effects of students' prior achievement and attendance, as well as the effects of school variables such as mobility and crowdedness, percentage of minority and free/lunch recipient students, and neighborhood census variables. The portion of achievement left over from this second stage model are again termed residuals. These residuals represent the portion of achievement that cannot be accounted for by variables schools do not control. They can be viewed as estimates of the schools' uncontaminated contribution to achievement, or school effectiveness. Teacher effectiveness is also measured in a similar manner. (Webster & Menro, 1997).

The School Improvement Indices ensure that the accountability system is valid and fair by not penalizing districts for factors, such as student SES, that are beyond their control. The Indices also measure the value that schools and the district provide to each student. Schools and their staff are eligible for financial awards based on school performance on the Improvement Indices.

Because of concerns that teachers might teach narrowly to specific tests, Dallas uses two different tests, the Stanford 9 and the TAAS. The Stanford 9 is not based on the Texas standards and is administered nationally. The TAAS is aligned with the Texas standards. The use of both tests minimizes the danger that teachers will, under pressure to raise scores, attempt to teach a narrow body of knowledge based on the likelihood of certain questions appearing on one type of test.

Tennessee

In the 1990s, the state legislature and governor stated that public education in the state of Tennessee was an impediment to industrial recruitment and economic development, and proposed the establishment of an accountability system. William Sanders, a University of Tennessee professor, was invited to make a presentation to the governor and legislature and, in 1992, his methodology was incorporated as the cornerstone of the Tennessee accountability system. Sanders designed the Tennessee Value-Added Assessment System (TVAAS), a statistical model that estimates the contributions of districts, schools, and teachers to student achievement.

TVAAS is based on the annual Tennessee Comprehensive Assessment Program given each spring in grades three through eight. The scaled score for each test in the TCAP

series increases with each grade level. With this scaling system, a student's score increase reflects his or her learning during that year. For example, the difference between a student's score on the third-grade math test and his or her score on the fourth-grade math test indicates how much the student learned in the fourth grade. The value-added reports complement the absolute achievement scores reflected in TCAP score reports and provide additional information on students' academic success. The state of Tennessee is committed to value-added accountability. According to the state Department of Education, "TVAAS assessments are based on the premise that schools and teachers have a significant role in student achievement and that it is possible to measure that effect by calculating the gains, or value-added, in student achievement (Tennessee Department of Education, 1997).

The TVAAS statistical model includes individual test scores and excludes the influence of variables such as the socioeconomic status of students. "Each child can be thought of as a 'blocking factor' that enables the estimation of school system, school, and teacher effects free of the socioeconomic confoundings" (Sanders, Saxton, & Horn, 1997). Each student's gains are compared against his or her own performance over three years. These gains are aggregated to provide school and district average gains. Tennessee then compares each school and district's gain with a national norm. Three-year average gains are used to determine whether systems are meeting the requirement of making 100 percent of the national norm groups' gains in the academic areas.

Tennessee's State Department of Education has established TVAAS-based performance standards for schools and districts. Minimum expectations are that the average student in a district will gain a year's average growth (compared to the national norm) for a year's instruction in each subject area and that the average TVAAS score for a school or district will be at the national average. If a district or school persistently fails to meet these standards, it can trigger state intervention. In addition to the TVAAS, Tennessee's accountability system includes an incentive award program based on student attendance rates and graduation rates. Teacher effectiveness is also measured using the value-added system. A teacher's effectiveness is determined by comparing the gain a teacher's students achieve during the year they are together with growth expectations based on the performance of the same students during the previous three years.

Texas

The original Texas Academic Excellence Indicator System or AEIS goes back to 1984, when the Texas Legislature for the first time sought to emphasize student achievement as the basis for accountability. That year, House Bill 72 called for a system of accountability based primarily on student performance. Prior to that, accountability focused mostly on process, that is, districts were checked to see if their schools had been following rules, regulations, and sound educational practices. Since the first year of the AEIS (1990-91), it has developed and evolved through legislative amendments, the recommendations of advisory committees and the commissioner of education, State Board of Education actions, and final development by Texas Education Agency researchers and analysts.

An assessment system based on the Texas Assessment of Academic Skills (TAAS) was implemented in 1990-91. TAAS is a criterion-referenced test based on the original 1985 Texas Essential Knowledge

and Skills tests. TAAS is administered in grades 3-8 and 10 in reading and mathematics, in grades 4, 8, and 10 in writing, and in grade 8 in social studies and science. Growth between years can be calculated for an individual student, a campus, a district, or the state. TAAS tests are administered in the late spring of each school year.

TAAS is designed to measure problem-solving and critical thinking skills required in the state-mandated curriculum, rather than minimum skills. The purpose of TAAS has expanded from the school-level diagnosis of individual student performance to its current use as a state-level evaluation tool to hold schools accountable for student performance (Texas Education Agency, 1997).

In 1993, Texas developed a rigorous accountability system to accredit school districts and rate schools. Texas Education Agency staff, educators and school board members, business and community representatives, professional organizations, and legislative representatives across the state collaborated on the system design.

Schools and districts are annually evaluated based on three indicators. TAAS performance, dropout rates, and attendance rates are used to determine district accreditation status and campus performance ratings. The TAAS performance indicators – the percentage of students passing each test (reading, writing, and mathematics) summed across grades - are evaluated for individual student groups (African American, Hispanic, White, and economically disadvantaged), as well as for all students tested.

The criterion for state rewards is based on the Campus Comparable Improvement (CCI) and Academic Excellence Indicator System (AEIS) school ratings. CCI is the value-added component of the school rating system. When a school rated as acceptable or above under the AEIS system also demonstrates significant performance gains in CCI performance, it will get rewards. Campus Comparable Improvement is measured by calculating academic gains by subject from year to year for individual students. These gains are then averaged to find the school's gain. These gains are then compared to schools that serve a similar student population. To find similar schools, the proportion of African American, Hispanic, white, economically disadvantaged, and limited English proficient students are compared. Longitudinal comparisons across years and across grades within a subject area for reading and mathematics at grade 3-8 and 12 can be made.

CCI compares the gains within each school group having similar characteristics over time. Each student's record is matched to his or her previous year's record. After matching student records, each student's gain is calculated. The reading and mathematics growth for students who were tested for two consecutive years in that school is then aggregated. This value-added gain score is used to measure student progress. Each school is then ranked relative to the forty Texas schools that are considered comparable to it.

North Carolina

The North Carolina State Board of Education developed the ABCs of Public Education in response to the School-Based Management and Accountability Program enacted by the General Assembly in June 1996. The ABCs program was a component

of a broad education reform effort led by Governor Hunt. The reform effort also included aligning assessment and curriculum, raising teacher salaries, improving professional development opportunities for teachers, and promoting school readiness. The ABCs focuses on strong **A**ccountability with an emphasis on high educational standards; teaching the **B**asics; and maximum local **C**ontrol. In 1998-99, a single comprehensive ABC model for elementary, middle and high schools was implemented. That model included measures of expected growth and minimum achievement standards. The state sets growth goals and performance standards in reading, writing, and mathematics at the elementary/middle school levels. The growth standards are school-based, and change with each year's cohort of students. The statewide performance standards are also school based.

North Carolina has created its own End-of-Course (EOC) test and End-of-Grade (EOG) tests; both are criterion-based. Each of the tests has four achievement levels with students expected to attain at least a level III. The results from these tests are used to calculate the school's overall performance composite. The test results are also used to compute a growth rate.

The status of schools and the incentive awards that schools receive are determined by the two components in the ABCs:

1) The performance composite tells the percentage of student test scores that are at or above Achievement Level III in the subjects tested.

2) The expected and exemplary growth/gain composite shows whether a school achieved the expected growth rate each year. *Positive* means that the school met or exceeded its expected growth goal. *Exemplary* indicates that a school achieved 10% more than the expected gain.

North Carolina puts a higher priority on the growth composite than the performance standard. Even if a school has more than 90 on its composite score, it will not be recognized as an excellent school unless the expected growth is achieved. Schools located in high income areas would be expected to have a high performance composite, but may not be recognized as excellent schools because students do not reach growth targets. On the other hand, a school that continues to show student growth may be recognized as an "expected growth school," even though the performance composite is less than 50.

North Carolina calculates the expected growth of cohort groups (the expected value-added of a teacher or school) based on state norms from the 1992-3 and 1993-4 school year. The expected growth rate is the average growth rate of the state's students in each grade between the 1992-3 and 1993-4 school year. For instance, the average reading score of North Carolina's third graders was 142.7 in the1992-3 school year. The average reading score for fourth graders was 147.9 in the1993-94 school year. The 5.2 point gain is the basis for expected reading growth from grade three to grade four. In the calculation of each school's expected growth rate, previous achievement level, regression to the mean, and students' proficiency are controlled (North Carolina Department of Public Instruction, 2000).

After determining the expected growth rate, the state standardizes the growth rate of each grade and each subject to determine a school's growth composite. If a school's standardized expected growth composite is positive or zero, then the school achieves expected growth. If a school exceeds the expected growth composite by more than 10%, the school will get exemplary recognition. Since this formula measures the cohort group's growth rate in math and reading from 3rd through 10th grade, a school can be held accountable for the achievement growth of cohort groups over time (North Carolina Department of Public Instruction,2000).

Section III: How Can Michigan Include Value-Added Assessment in the State's Accountability Model?

Michigan has already put into place the sophisticated Michigan Education Information System (MEIS) that provides an infrastructure for many of the necessary data elements for a fair and effective accountability system. The Single Record Student Database (SRSD) contains the necessary elements to begin a tracking/accountability system that is based on individual student information. Every student in the state has been given a unique identifier based on the first and last name of the child, the date of birth, and gender. Using the Single Record Student Database, the state can append additional information about student achievement in order to create a comprehensive accountability system. But what else should the state do?

- ***Implement annual testing***

Congress has mandated annual testing for all students in grades 3 through 8. Annual testing would provide two important indicators for accountability: *the annual gain* in test scores from one year to the next, i.e. value-added by a teacher and school, and *the mastery of standards*. Each year's test scores can be compared to previous scores by using national norms or state standards, whichever is applicable. Comparisons can be made between classes at specified grade levels, among buildings within a district, or among districts. Data can also be disaggregated by subgroups such as students eligible for free or reduced-price lunch, or students whose native language is not English or race/ethnicity. This would allow districts to better identify students who are not progressing. Once the program has been in place for a while, annual tests would also give schools multiple years of assessment data to more fairly measure changes in student achievement.

To implement annual testing, Michigan must decide whether to purchase an "off the shelf" norm-referenced test, such as the Stanford 9 or the Iowa Test of Basic Skills, or to construct its own tests based on the Michigan Core Curriculum Framework. Norm-referenced tests are relatively inexpensive and will make it easy to compare Michigan students with students in other states or across the nation. Unfortunately, the tests will not be aligned with the Michigan Core Curriculum Framework. On the other hand, if the state chooses to develop it own annual tests, they are more likely to be aligned with the standards and curriculum of the Michigan Core Curriculum Framework but test development can be an expensive and complicated process.

What tests are other states using?

Tennessee administers a comprehensive set of nationally normed, criterion based achievement tests annually to students in grades three through eight. The tests are produced by CTB (CTB-McGraw/Hill) and are similar to TerraNova, but customized to Tennessee. Tennessee teachers are involved in item development and test items must be approved by teachers and by staff of the Tennessee Department of Education assessment division.

The Tennessee Comprehensive Assessment Program (TCAP) measures attainment levels as well as the annual progress of students. Students are tested in grades three through eight in five subject areas: math, reading, language arts, science, and social studies. A writing test is required for elementary and middle school students. These tests provide norm-referenced data for national comparisons as well as criterion-referenced information for use in determining whether students have mastered specific state instructional objectives. These tests also provide the data necessary for TVAAS.

Texas assesses students using the Texas Assessment of Academic Skills (TAAS). This criterion referenced test was developed by a collaborative group of State Board of Education personnel, classroom teachers, administrators, and curriculum specialists. Since the program began in 1990 over 6000 teachers have served on test development committees. The tests are released to the public at the end of each annual testing cycle. While the tests are criterion referenced, each annual revision of the test is benchmarked before the passing score is established.

Beginning in grade three, students participate annually in reading and mathematics tests throughout elementary and middle school. They are also tested once in high school. Writing assessments are administered in grades four, eight, and ten. Science and social studies knowledge is assessed one time each during elementary, middle, and high school.

North Carolina implemented a statewide testing program in 1992-93 with tests designed by North Carolina teachers, curriculum specialists, testing experts, and the Department of Public Instruction staff. The testing program was modified to focus on the basics of reading, mathematics, and writing under North Carolina's ABCs of Public Education.

- ***Develop a value-added assessment model to measure school effectiveness***

Michigan can either draw from experiences in other states to develop its own value-added assessment model or it can replicate what has been done elsewhere. The choices include:

(a) Adopt a mixed model application similar to the Tennessee approach

Models such as the one adopted by Tennessee require complex software that runs a mixed model application to measure teacher effectiveness by using test scores. Michigan could adopt a nationally norm based test such as the SAT9 or develop a set of tests similar to Tennessee's nationally normed tests. Once these tests are in place, the state could either contract with Sanders' group (the one used in Tennessee) or contract out the development of a new model based on the criteria established by Sanders' group.

The advantage of this option is that the Tennessee model provides teachers and schools with detailed information regarding their effectiveness. However, this model is based on the assumption that students are assigned to teachers randomly. This assumption could result in inaccurate measurement of teacher effectiveness. The other major disadvantage of this type of approach is that Sanders has not shared his model with others. He has treated the model as proprietary information and has not allowed the general public or the research community to examine his methodology or to independently verify his findings. At the school and system level, Tennessee uses a fixed effect ANOVA model and does not include SES variables in the model. At the teacher level, Tennessee uses a mixed ANOVA model that also does not include SES variables. Sanders argues that longitudinal analysis adequately controls for students characteristics that do not change over time. The validity of this argument is still being debated within the research community. (For more detail on the Tennessee model, see Millman, J. (ed). 1997 "Grading Teachers, Grading Schools, Corwin Press)

(b) Adopt a regression approach similar to the DISD model

The DISD model is a two stage model using both regression techniques and a hierarchical linear model (HLM). This model is often referred to as a residual analysis. In the first stage, Dallas controls for mobility, percentage minority, percentage free-lunch eligibility, and other student variables that are outside of the school's control by using a multiple regression model. Residuals from this first stage regression model are used in the second stage HLM model which controls for prior achievement and teacher, school, and neighborhood variables that are outside of educators' control. The residuals obtained from the second stage HLM model are then used to measure school and teacher effectiveness.

Like the Tennessee model, this model is difficult for the public to understand. However, it has some advantages when compared to the Tennessee approach. The first is that education researchers are more familiar with this type of regression model than with the Tennessee model. The model is not treated as proprietary information so independent researchers have been able to test the claims made by proponents of the system. In addition, since the model filters out variables outside of the schools control (referred as fairness variables in the DISD), it can assuage the concerns of legislators, educators and the public about the equitable measurement of school effectiveness.

c) Adopt an algebraic growth measurement similar to the NC or Texas models

North Carolina and Texas use relatively simple models to measure growth in student achievement. Texas measures achievement gains by using equated scores of matched students and aggregating them to the school level. This gain is then compared to the gains made by schools that serve similar student populations.

North Carolina computes growth gain by using a simple algebraic model and adjusting for regression to the mean and true proficiency. The North Carolina model does not control for student race or socioeconomic background. The benchmark

of each school's expected growth rate is based on the average growth rate of North Carolina students in the 1992-93 and 1993-94 school years. This benchmark is then adjusted by indices for true proficiency and regression to the mean and for coefficient of estimation.

The public generally finds these algebraic growth models easier to understand but this simplicity comes with a price tag. These models do not completely account for factors such as student background that the school is unable to alter but that do affect student achievement, and this may undermine public and professional confidence in the fairness of the accountability system.

- ***Develop Incentives and Supports based on the Value-Added Assessment Results***

The most difficult part of an accountability system is designing and implementing a system of incentives and supports: incentives for schools that improve student achievement and support for schools where students fail to make progress. Rewards should be based on indicators that are fair and valid, such as three years of value-added testing data in combination with other indicators of success, such as dropout and attendance rates.

An important goal for value-added assessment is to encourage teachers and schools to enable all children to succeed and to continually progress. To accomplish this, an accountability system should include some type of incentive or compensation for schools where children continue to achieve high standards, and also for schools where children progress academically. In order to be fair, and to insure that schools in high-income areas don't receive a disproportionate share of rewards, both measures need to be in place. The value-added component levels the playing field, holding schools responsible only for those factors under their control. High standards provide an incentive for all schools, even those with challenging demographics, to push for high absolute levels of academic performance.

The creation of rewards should be continued for a number of years. The appropriation of funds must be sustained over an extended time if teachers and administrators are to view rewards as useful and valuable. In addition to school level rewards, North Carolina and Texas provide incentive pay to teachers whose schools accomplish expected gains. School level incentive programs have become more popular than individual merit pay for teachers, since this encourages cooperation among teachers to improve overall student learning within a school.

Currently, Michigan administers a school-level incentive award, the Golden Apple Award, based on MEAP test scores. The state may want to consider expanding the current Golden Apple Award based on the value-added assessment system.

Conclusion: Where Do We Go From Here?

Michigan has recognized that its current accountability system is inadequate. The state has begun the process of developing an accountability system that includes the requisite parts – the construction of a Single Record Student Database and MEIS, the development of an annual testing program, the use of multiple indicators in evaluating student progress, and instituting a set of rewards for schools where students meet state goals for progress.

While there is no one best system of accountability, the fairness of the system must be a key concern. When educators perceive a system as fair, they are much more likely to accept and support the system (Elmore, Abelmann, & Fuhrman, 1996). When teachers are provided with fair assessments of their students' achievement, they can use the information in a thoughtful manner to reflect on and improve their instruction. An important component of any fair accountability system is the inclusion of a value-added testing program. Value-added testing responds to the state superintendent's request that, before implementing any new program, we ask ourselves, "What does this do to help teachers teach and students learn?" By providing teachers, parents, legislators, and the general public with a fair measure of student learning, value-added testing helps to ensure that the accountability system is a tool for improved teaching and learning.

References:

21st century report card(1997). Tennessee Department of Education. Available: http://www.k-12.state.tn.us/arc/rptcrd97/tvaas.htm [2001, December 19].

Bryk, A. S., Deabster, P. E., Easton, J. Q., Luppescu, S., & Thum, Y. M. (1994). Measuring achievement gains in the Chicago public schools. *Education and Urban Society*, *26*(3), 306-319.

Cooley. (1991). State-wide student assessment. *Educational Measurement: Issues and Practice*.

Determining composite scores in the ABCs model: 1999-2000(2000). North Carolina Department of Public Instruction. Available: http://www.dpi.state.nc.us/accountability/reporting/2000Memo/abc2000.PDF [December 19, 2001].

Elmore, R. F., Abelmann, C. H., & Fuhrman, S. H. (1996). The new accountability in state education reform: From process to performance. In H. F. Ladd (Ed.), *Holding schools accountable: Performance-based reform in education* (pp. 65-98). Washington, DC: Brookings Institution.

Expanding the scope of the Texas public schools accountability system(1997). Texas Education Agency. Available: http://www.tea.state.tx.us/research/pdfs/prr9.pdf [2001, December 19].

Hanushek, E. A. (1994). *Making schools work*. Washington, D.C.: Brookings Institution.

Hanushek, E. A., & Meyer, R. H. (1996). Comments on chapter two, three, and four. In H. F. Ladd (Ed.), *Holding schools accountable: Performance-based reform in education* (pp. 128-145). Washington, DC: Brookings Institution.

Kane, T. J., & Staiger, D. O. (2001). *Improving school accountability measures*. Cambridge, MA: National Bureau of Economic Research.

King, R. A., & Mathers, J. K. (1997). Improving schools through performance based accountability

and finance rewards. *Journal of Education Finance, 23*, 147-176.

Ladd, H. F. (1999). The Dallas school accountability and incentive program: An evaluation of its impacts on student outcomes. *Economics of Education Review, 18*, 1-16.

Ladd, H. F., & Clotfelter, C. T. (1996). Recognizing and rewarding success in public schools. In H. F. Ladd (Ed.), *Holding schools accountable*. Washington, DC: Brookings Institution.Mathers, J. (1999). *Performance-based accountability systems.* Education Commission of the States. Available: http://www.ecs.org/ecsmain.asp?page=/search/default.asp [2001, December 19].

Meyer, R. H. (1997). Value-added indicators of school performance: A primer. *Economics of Education Review, 16*(3), 283-301.

Quality counts 99 (1999).: Education Week.

Raudenbush, S. W., & Willms, J. D. (Eds.). (1991). *Schools, classrooms, and pupils*. San Diego, CA: Academic Press.

Sanders, W. L., Saxton, A. M., & Horn, S. P. (1997). The Tennessee value-added assessment system: A quantitative, outcomes-based approach to educational assessment. In J. Millman (Ed.), *Grading teachers, grading schools: Is student achievement a valid evaluation measure?* (pp. 137-162). Thousand Oaks, CA: Corwin Press.

Setting annual growth standards: "The formula"(2000). North Carolina Department of Public Instruction. Available: http://www.ncpublicschools.org/accountability/reporting/asb_formula.pdf [2001, December 19].

Taylor, T. (1985). A value-added student assessment model: Northeast Missouri state university. *Economics of Education Review, 4*(4), 341-350.

Webster, W. J., & Mendro, R. L. (1995). Evaluation for improved school level decision-making and productivity. *Studies in Educational Evaluation, 21*, 361-399.

Webster, W. J., & Menro, R. L. (1997). The Dallas value-added accountability system. In J. Millman (Ed.), *Grading teachers, grading schools: Is student achievement a valid evaluation measure?* (pp. 81-99). Thousand Oaks, CA: Corwin Press.

Webster, W. J., & Olson, G. H. (1988). A quantitative procedure for the identification of effective schools. *Journal of Experimental Education, 56*, 213-219.

Websites for further information:

Michigan Department of Education
http://www.meritaward.state.mi.us/
http://cepi.state.mi.us/

North Carolina Department of Public Instruction, Determining Composite Scores in the ABCs Model, 1999-2000, http://www.dpi.state.nc.us/accountability/reporting/2000Memo/abc2000.PDF

North Carolina Department of Public Instruction, Setting Annual Growth Standards: The Formula, 2000, http://www.ncpublicschools.org/accountability/reporting/asb_formula.pdf

North Carolina Department of Public Instruction, 1999-2000 ABCs Recognition Categories, http://www.ncpublicschools.org/abcs/Categories.html

North Carolina Department of Public Instruction, The End-Of-Course Prediction Formulas, http://www.ncpublicschools.org/accountability/reporting/Prediction_Formula/pred13Sep00.html

Tennessee Department of Education, 21st Century Report Card: Tennessee Value-Added Assessment System, http://www.k-12.state.tn.us/arc/rptcrd97/tvaas.htm

Tennessee Department of Education, How to Interpret the Report Card, http://www.k-12.state.tn.us/rptcrd00/rptcrd.htm

Texas Education Agency, 2001 Accountability Manual, http://www.tea.state.tx.us/perfreport/account/2001/manual/index.html

Texas Education Agency, 1999-2000 Technical Digest, http://www.tea.state.tx.us/student.assessment/resources/techdig/index.html

Expanding the Scope of the Texas Public School Accountability System, a Policy Research Report of Texas Education Agency http://www.tea.state.tx.us/research/pdfs/prr9.pdf

Website of SAS in School inc. http://www.sasinschool.com/evaas/index.shtml

Website for CPRE http://www.gse.upenn.edu/cpre/Publications/nc.pdf

The Education Policy Center at Michigan State University
201 Erickson Hall , Easting Lansing, Michigan 48824-1034
www.epc.msu.edu