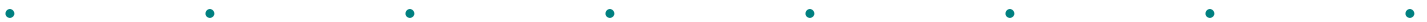




Value-added Assessment Systems



November 2004



Value-Added Assessment Systems

Introduction

“Value-added,” a term originally used in business and economics, has become widely used to describe certain educational assessment and accountability systems. The most widely known and frequently cited value-added system is the Tennessee Value-Added Assessment System (TVAAS) developed by William Sanders and in use since 1996 (McCaffrey, Lockwood, Koretz, and Hamilton, 2003). TVAAS has since been credited with leading to the implementation of value-added assessment systems in states, districts, and consortia nationwide (Carey 2004; Hershberg, Simon, and Lea-Kruger 2004; Kupermintz, 2003). However, TVAAS is only one of many existing value-added models. As states begin considering the adoption of value-added systems, policymakers and education stakeholders must become more familiar with this new approach. This report provides an introduction to value added assessment for policy makers and stakeholders by discussing its fundamental concepts, examining its potential applications in education systems, and reviewing its most pressing issues and challenges.

Fundamental Concepts of Value-Added Assessment Systems

“Value-added assessment system” does not refer to one particular test format. Rather, value-added refers to any one of several models that are used to interpret test scores in a way that evaluates the growth or progress in a student’s academic achievement over time, usually over several academic years (Rubin, Stuart, and Zanutto, 2004). The concept of an assessment that measures a student’s achievement growth over several years, commonly known as *longitudinal assessment*, has long existed in education (Goldschmidt, Choi, and Martinez, 2004). However, value-added assessment represents an approach to evaluating student achievement growth that is distinct from traditional growth models in several respects.

Value-Added Measurements versus Simple Growth Scores

Value-added models are frequently compared to growth scores, which have long been used in educational assessment. A growth score is typically calculated as the difference between a student's scores for the current year and the previous year. Value-added models differ markedly from simple growth scores calculations through the use of sophisticated statistical formulas that are intended to isolate noneducational factors, such as the student's socio-economic status and demographics. Once these factors are isolated, their impact is removed from the measure of the student's achievement growth. With noneducational factors removed, the unadulterated measure that is produced is considered to express the student's true achievement growth. More importantly, the value-added score is interpreted as a measure of the direct effect that educational factors—the teacher, school, or district—have had on the student's achievement (Drury and Doran, 2004; Hershberg et al., 2004; McCaffrey et al., 2003).

Determining Effective Teachers and Schools

A fundamental concept of value-added assessment is the assertion that schools are responsible for providing each student with the equivalent of one year of growth, regardless of the level of education with which the student begins the academic year (Callender, 2004; Carey, 2004; Hershberg et al., 2004). If the value-added measure reflects the true effect of teachers and schools on a student, it should be possible to determine whether their contribution to the student's growth was sufficient.

Each value-added model formulates what constitutes a year of growth for a student. In the case of TVAAS, a year of growth is estimated using each student's individual achievement history (Bianchi, 2003; Hershberg et al., 2004). At the end of the year, if the student has achieved what was computed to be one year of growth, the student is considered to have received an effective education. If the student shows more than one year of growth, the student has received a highly effective education. Students that show less than one year of growth are considered to have received a less effective education (Hershberg et al., 2004).

The quality of the student's education is considered to have long term implications. Relying on the findings from independent studies of TVAAS, supporters of value-added assessment claim that teacher effectiveness is far more important in a student's current and future achievement than any other non-educational factor (Bianchi, 2003; McCaffrey et al., 2003). They assert that, at later grade levels, students who receive less effective educations will continue to demonstrate low levels of achievement while students who receive effective educations will attain higher levels of achievement. Some educational researchers



point to value-added assessment as supporting the view that teachers play the dominant role in student achievement growth (Vaughan 2002).

Requirements for a Value-Added Assessment System

Another difference of value-added assessment systems from traditional growth scores are the resources required. For an education system to implement a value-added assessment system, several considerations must be made:

- There must be a system for tracking the achievement scores and demographic information of each student over time.
- The data must be stored in a database capable of handling a state-sized student population.
- Computing resources must be available to handle the complex statistical calculations that often will involve the achievement histories of the student population of an entire district or state.

While value-added assessment is independent of any particular achievement test, there is agreement that the scores used in the analysis must be produced by a reliable and valid achievement test that is administered at least annually and can express scores on a common scale that is meaningful across grade levels (Stone, J. E., 1999; Meyer, 1996; Goldshmidt, Choi, Martinez, 2004; Herdman et al., 2002; Drury and Doran, 2003). That is, the assessment must produce *scaled scores*, which place a student's scores on a vertical scale that corresponds to an achievement continuum and can show progress over time (Nitko, 2004). In a recent study of value-added models, researchers relied on scaled scores from the *Stanford Achievement Test Series, Ninth Edition* (McCaffrey, Koretz, Louis, and Hamilton, 2004). Research by Goldshmidt et al. (2004) used the same assessment in a study that demonstrated that the use of scaled scores in value-added formulas provides more consistent results than the use of other measures, such as percentile ranks and normal curve equivalents (NCE).

Applications of Value-Added Assessment Systems

Proponents of value-added assessment believe that this new kind of data will make it possible to evaluate and compare the quality of schools that have widely different student populations (Meyer, 1996). This assertion is a marked departure from conclusions drawn from studies, particularly the well-known "Coleman Report", that claimed to show that a student's socioeconomic status and demographics, rather than teachers and schools, have the most impact on his or her achievement (Bracey, 2004). Since the publication of those studies, the evaluation of schools has typically relied on measures of educational inputs, such as funding and teacher certifications, rather than test results (Meyer, 1996).



However, in the new era of accountability led by the *No Child Left Behind Act of 2001* (NCLB), educational research has supported the view that instruction has a real impact on students that can be measured with standardized assessments (Archer, 1999; Marzano, Pickering, and Pollock, 2001). Such conclusions support anecdotal evidence from parents and other stakeholders who observe a wide range of teacher effectiveness (McCaffrey et al., 2003).

Using Value-Added Systems for Educational Improvement

Commentators foresee potential for using a value-added assessment system to guide large-scale and small-scale educational reform. The principle expectation for the value-added model is that the results will conclusively determine the impact that educators and education policies are having on their students (Drury and Doran, 2004; McCaffrey et al., 2003). With this evidence, effective teachers and education policies can be identified, and reform based on these findings can be instituted in education systems (Carey, 2004; Crane, 2002). For example, results from a value-added system could be used to transfer effective teachers to schools where they are needed (through financial incentives or other means), study the instructional practices of effective teachers, and offer professional development to less effective teachers (Carey, 2004; Drury and Doran, 2004; Hershberg et al., 2004). Other, more controversial, suggestions include holding educators accountable for student growth, such that highly effective teachers receive financial incentives and professional advancement while consistently ineffective teachers are sanctioned (Summers, 2002).

Accountability systems based on value-added assessment have already been instituted in some states. For example, school districts in Chattanooga, Tennessee use TVAAS to identify and attract highly effective teachers using salary bonuses, housing benefits, and funding for graduate education (Carey, 2004). As most education systems rely on seniority to determine teacher transfers and salary schedules, implementing a reward or evaluation system based on value-added assessment data may represent a significant change in policy (Carey, 2004; Drury and Doran, 2004).

NCLB and AYP

Another potential application of value-added assessment is a role in the implementation of NCLB. The accountability mandate of NCLB requires states to demonstrate, through annual assessments, that schools are making Adequate Yearly Progress (AYP) in the achievement of their students. Hence, achievement tests used for accountability are already expected to indicate the contribution that a school is making to a student's educational growth.



Supporters of value-added assessment observe that, typically, an analysis of AYP is made using a *cohort comparison*. In this methodology, the scores for the current students in a certain grade are compared to the scores of the previous year's students in the same grade (Choi, Seltzer, Herman, and Yamashiro, 2004; Crane, 2002; Hershberg et al., 2004; McCaffrey et al., 2003). Cohort comparisons show progress at the school level using “snap-shots” of completely different sets of students with little to explain the increase or decrease in achievement (Drury and Doran, 2003). This kind of analysis is also known as a *level-indicator* as it determines the quality of a school by whether average student achievement is above a certain threshold or level (Meyer, 1996).

However, the use of level indicators has received the criticism that it “confuses the school building for the students” by diluting the achievement of individual students (Wenning, Herdman, and Smith, 2003). A cohort-comparison relies only on the design of the standardized test to account for the non-educational factors that can affect student scores (Herdman, Smith, and Doran, 2002). Some educational researchers suspect that level indicators, lacking additional controls for bias, produce misleading assessments of schools with large numbers of academically disadvantaged students (Meyer, 1996). Moreover, when a cohort-comparison is used, teachers may focus only on students who will make it possible for a school to reach AYP, potentially ignoring the educational progress of the highest and lowest achievers (Hershberg et al., 2004; Meyer, 1996). Critics suggest that AYP puts schools with large numbers of disadvantaged students at an unfair disadvantage, characterizing this shortcoming as “tantamount to measuring a child’s height with a yardstick but acknowledging growth only when his or her height exceeds 36 inches” (Drury and Doran, 2004, p. 3).

In contrast to a cohort-comparison, value-added assessment emphasizes the achievement growth in individual students over time, an approach that may be perceived as closer to the spirit of NCLB (Callender, 2004; Herdman et al., 2002). Criticism of AYP as it is currently practiced arises from the suspicion that some schools make great progress with disadvantaged students but still fail to reach AYP while other schools with talented and advantaged students reach AYP without providing an effective education (Herdman et al., 2002; Stone, 1999). Researchers have recently lent credence to this suspicion, having found in simulations that a significant number of schools that met AYP were determined to be ineffective by value-added analysis. In other cases a “remarkably large numbers of non-AYP schools” were considered effective by value-added analysis (Choi et al., 2004, p. 9). Proponents of value-added assessment assert that their model recognizes the reality that students start at different levels of achievement and that a school can be considered effective when it provides one year of growth to even the lowest-achieving students (Hershberg et al., 2004).



By using a value-added assessment system to identify the classrooms with ineffective teachers, schools and districts may also predict which schools may be at risk of not reaching AYP (Hershberg et al., 2004). Assistance or intervention can then be provided where it is needed. By placing low-achieving students with highly-effective teachers and schools, districts will be able to work efficiently towards closing the achievement gap and bringing all students to proficiency, thereby realizing the promise of NCLB (Drury and Doran, 2004; Hershberg et al., 2004). Hence, value-added assessment may be seen as a complement to AYP by determining whether low achieving students are moving towards proficiency at a greater rate than expected (Carey, 2004).

Others suggest that value-added assessment should be a component or at least a supplement of each state's definition of AYP (Choi et al., 2004; Herdman, Smith, Doran, 2002). The U.S. Department of Education, in draft regulations and other communications to policymakers, has encouraged the development of alternate methods for determining AYP, especially growth-based measures such as value-added assessment (Doran, 2003; Herdman et al., 2002). However, value-added assessment systems have not yet been approved by the federal government for use as AYP models in states.

Issues and Challenges to Value-Added Assessment Systems

While value-added assessment clearly has great potential to have a positive impact on education systems, some commentators have reservations about a number of issues. The most commonly cited concerns include the statistical methods used to account for bias from non-educational factors, some fundamental issues of scientific research, the effect of missing or erroneous data, the transparency of the statistical model to stakeholders, the limited subject areas that can be evaluated, and questions about the available published research.

Bias Caused by Non-Educational Factors

Perhaps the most significant feature attributed to value-added assessment models is the removal of non-educational factors, such as socioeconomic status and demographics, so that measured growth in a student's learning can be attributed directly to the teacher or education system. The statistical methods by which various value-added models accomplish this ambitious goal have been the subject of extensive scrutiny.

In the TVAAS model, the student's own history of test scores is used as a point of comparison to the student's measured growth, an approach known as "blocking" (Bianchi, 2003; Kupermintz, 2003). Proponents of this approach assert that the model accounts for not only the bias introduced by a student's race or



socioeconomic background but also for the bias commonly held to be inherent in standardized test scores (Ballou, Sanders, and Wright, 2004; Carey, 2004; Hershberg et al., 2004;). A team of RAND researchers who studied this method remains unconvinced that this approach satisfactorily accounts for bias, warning that noneducational effects may be attributed mistakenly to teachers with no way of determining the magnitude of this error (McCaffrey et al., 2003). Examining available data, Ballou (2002) and Kupermintz (2003) suggest that non-educational factors have a noticeable impact on the evaluation of teachers despite attempts to control these variables with blocking.

In another approach to address bias, the Dallas Value Added Accountability System uses a statistical formula which incorporates variables representing non-educational factors, including ethnicity, language proficiency, gender, and free-or-reduced-lunch status. Some commentators object to a value-added system that includes demographic information in its calculation on the grounds that a lower expectation will be set for students who are disadvantaged or from certain ethnic groups (Meyer, 1996; Olson, 1998). However, Drury and Doran (2003) suggest that schools with disadvantaged students are likely to have more ambitious value-added performance goals than other schools. Combined with the NCLB-mandated high achievement standards for each student, a value-added assessment system may be seen as raising the expectations for disadvantaged students. Value-added assessment puts gains in perspective, but does not lower the ultimate goal of high standards for all students.

Another non-educational factor to consider is the possibility that a student's aptitude could play a role in his or her own learning growth. Some commentators assert that a student's growth in response to effective instruction depends on the aptitude of the student. That is, teachers who are assigned to more challenging students are likely to be evaluated as less effective than teachers who are equally effective but responsible for students with high aptitudes for learning (Kupermintz, 2003). The effect that students could have on one another through their own independent interactions, such as informal tutoring, could also confound attempts to connect a teacher to learning growth (Rubin et al., 2004).

Some critics of value-added assessment contend that determining the effect that a teacher has on a student's learning growth is impractical. They suggest that there are too many factors, too much complexity, too many sources of bias, than can ever be expressed in a statistical formula. Kupermintz (2003) points out the familiar example of multiple teachers in different subjects having an impact on a student's achievement.

When a science teacher emphasizes the computational aspects of the curriculum and requires his students to engage in intensive mathematical explorations, increased student mathematical proficiency should be expected. When the math teacher collaborates or coordinates her efforts with the science teacher to help students meet the elevated demands of the science curriculum, further facilitation of students' math ability may be realized. . . .Attempts to disentangle such complex, interwoven contributions of the science teacher, the math teacher, and the computerized learning environment into isolated, independent "effects" are not only methodologically intractable but also conceptually misguided. (p. 290)

This criticism addresses the fundamental concepts of value-added assessment. If one flatly denies that statistical models can accurately measure the effects of education systems in a useful way, there is little more to say. While it may be difficult to understand how a formula could make it possible to connect a teacher to his or her classroom's learning growth, a value-added model must somehow account for these factors (McCaffrey et al., 2003).

Scientifically Linking Teacher Effects to Student Growth

Closely related to concerns about bias is an important, but more nuanced, criticism about the scientific aspect of value-added assessment (McCaffrey et al., 2003; and Kupermintz, 2003). For a teacher to be identified as having had an effect on a student's educational growth, it must be feasible that there was some plausible alternative in which the student would have experienced greater or less growth, such as another teacher, school, or district. However, it does not seem clear that there is a justification for choosing one possible alternative over another as a point of comparison. Thus, the data does not tell us which specific practices make a teacher's instruction more effective at providing students with achievement growth. To address this issue, some value-added models use a statistically "average" teacher from the district as a point of comparison; other models fail to address this point at all (McCaffrey et al., 2003).

Along the same lines, the data used in value-added assessment is considered observational, not random and experimental. This challenge to value-added assessment is common to all educational research: randomization is the "gold standard" of scientific research into causal effects (Rubin et al., 2004). Relying on observational data, rather than randomized data, causes two problems for value-added models: observed teacher effects on student learning may be confused with unobserved, non-educational factors that effect student learning (the bias issue already discussed), and data that can link the achievement of individual students to individual teachers may be unavailable (McCaffrey et al., 2003). However,



using a randomized, experimental study (for example, by randomly assigning students to schools or teachers) to determine whether a value-added assessment system measures teacher effects is an unfeasible prospect for the U.S. education system (Kupermintz, 2002; Shavelson and Towne, 2002). For their findings to be considered rigorous, researchers must address this concern in some fashion.

Missing and Erroneous Data

As discussed above, missing and erroneous data on student achievement records represent a significant challenge to value-added assessment in two respects (Meyer, 1996; Rubin et al., 2004). Districts with a highly-mobile student population are faced with incomplete records of achievement scores for individual students, and newly-hired teachers do not have a record of student test scores by which their effectiveness can be evaluated (Meyer, 1996).

A robust value-added model must use a statistical methodology to account for either situation. However, doing so involves tradeoffs that make it very difficult to avoid misattributing effects, whether to teachers or to noneducational factors (McCaffrey et al., 2003). For example, TVAAS assumes that missing data for teachers and students is the same as the average for the district (Carey, 2004). Kupermintz (2003) questions this approach, suggesting that using averages introduces bias for teachers with small numbers of students or significant amounts of missing data. As a result of shrinking scores towards the average, effective teachers may be undervalued while an ineffective teacher might escape notice. Moreover, using averages to estimate missing data could result in different evaluations for two equally effective teachers at different schools. Such a result might be perceived as relying on assumptions about the schools rather than on actual data. Educational stakeholders may not accept such an approach, especially if the results are to be used in high-stakes decisions (Kupermintz, 2002).

Other value-added models address this problem of by excluding students with missing data from analysis. McCaffrey et al. (2003) note that because low-achieving students are more likely to be missing data, a value-added assessment that relies on this approach is likely to have significant bias and, therefore, questionable validity.

The Transparency Issue

A fundamental challenge to value-added assessment, and to standardized testing in general, is the error inherent in the statistical process used to assign scores (Crane, 2002). While anyone who accepts the use of standardized tests must also accept that a certain amount of error is unavoidable (Nitko, 2004), Ballou (2002) suggests that the error will be compounded when achievement scores are used



together to measure gains. To solve such statistical problems with assessment scores, value-added models rely on sophisticated formulas.

Critics of value-added assessment suggest that a methodology that relies on esoteric formulas is too complex for teachers, parents, and other stakeholders to understand. To qualify for use in educational evaluation, they argue, the formulas of a value-added model must be transparent to anyone who holds a stake in its analysis (Ballou, 2002; Drury and Doran, 2004; Hershberg et al., 2004; Meyer, 1996; Olson, 1998). This concern is clearly not one to be dismissed out of hand. However, supporters of value-added assessment offer the rejoinder that complexity does not automatically disqualify something from being practically useful, citing such examples as automobiles, computers, and medical equipment as complex, useful things with inner workings that are not understood by most people (Herdman et al., 2002; Hershberg et al., 2004; Summers, 2002). Nevertheless, the evaluation of a teacher can be significantly effected by the choice of one value-added model over another or by changes to the methodology within a model. Hence, a model's statistical methodology should be explained as clearly as possible for all stakeholders (Ballou, 2002).

Limitation of Subject Areas Suitable for Evaluation

NCLB requires annual standardized testing for grades 3 through 8, enabling schools to gather the achievement scores that are necessary to make a value-added assessment of a student's progress. However, schools are only required to test in the core subject areas of reading and English language arts, mathematics, and science (Andrejko, 2004). Unless states establish vast testing programs that include all grade levels and subjects, such as art and music, a large number of teachers cannot be included in evaluation systems that are based on the results of a value-added assessment system (Stone, D. E., 1999). Hence, the use of value-added assessment systems for evaluative purposes may be of limited value for some education systems.

Questions about Supporting Research

A 2003 study by a team of researchers at the RAND Corporation examined the research supporting several value-added models (McCaffrey et al., 2003). The RAND team notes that the claims made by proponents of value-added assessment are based on a "relatively small number of papers—several of them not published in the peer-reviewed literature. . ." (McCaffrey et al., 2003, p. xiii). Having analyzed the published and unpublished studies that have been cited in support of value-added assessment, and, in some cases, having replicated the research, the RAND team calls into question some of the conclusions drawn by the original studies. Their findings do not invalidate the studies, but rather suggest that the value-added models examined may have overstated the magnitude of teacher

effects while downplaying the role of non-educational variables. However, it is important to note that the RAND team, as well as other commentators, frames questions about the empirical evidence that supports value-added assessment as an opportunity to advance and strengthen value-added models through further research (McCaffrey et al., 2004).

Conclusion

Despite the issues faced by value-added assessment, there is reason to be optimistic about its potential. While the RAND team concludes that “the research base is currently insufficient . . . to recommend the use of [value-added models] for high-stakes decisions,” they concede that “as policymakers evaluate alternative models for school or teacher accountability, [the value-added model] should be given serious consideration even in light of its limitations” (McCaffrey et al., 2003, p. 119–120). These observations have led some to conclude that while value-added assessment can be useful for informing evaluations of teachers, it should not be the lone deciding factor (Ballou, 2002; McCaffrey et al., 2003).

Research and discussion of value-added assessment is ongoing and will lead to a better understanding of education systems. The possibility of creating an assessment that allows us to improve educational practices is too great an opportunity to dismiss. All participants in the nation’s education system have a role to play in determining whether value-added assessment will become a fixture of education policy. Test publishers can contribute by producing valid, reliable instruments with vertical scaled scores, such as Harcourt Assessment’s Stanford Scale; by offering expertise in psychometrics and assessment to parties interested in pursuing value-added assessment; and by developing models that emphasize the strengths and overcome the challenges. Educational stakeholders, from parents to teachers to students, can only benefit from further research, discussion, and refinement of value-added assessment systems.

References

- Andrejko, L. (2004, Spring). Value-added assessment: A view from a practitioner. *Journal of educational and behavioral statistics*, 29(1), pp. 7–9.
- Archer, J. (1999, May 5). Sanders 101. *Education week*, 18(34). Retrieved from <http://www.edweek.org/ew/vol-18/34sander.h18> on May 20, 2004.
- Ballou, D. (2002, Summer). Sizing up test scores. *Education next*. Retrieved from <http://www.educationnext.org/2002/16.html> on March 31, 2002.

- Ballou D., Sanders W., & Wright P. (2004, Spring) Controlling for student background in value-added assessment of teachers. *Journal of educational and behavioral statistics*, 29(1), p. 37–65.
- Bianchi, A. B. (2003, June). A new look at accountability: “Value-added” assessment. *Forecast: Emerging issues in public education*, 1(1), p. 1–4.
- Bracey, G. W. (2004, May). Serious questions about the Tennessee value-added assessment system. *Phi delta kappan*, 85(9), pp. 716-717.
- Callender, J. (2004, Spring). Value-added student assessment. *Journal of educational and behavioral statistics*, 29(1), p. 5.
- Carey, K. (2004, Winter). The real value of teachers: Using new information about teacher effectiveness to close the achievement gap. *Thinking K–16*, 8(1), pp. 1–42.
- Choi, K., Seltzer, M., Herman, J., & Yamashiro, K. (2004). *Children left behind in an AYP school?! Validation of AYP by focusing on student progress and the distribution of student gains*. Paper presented at the Council of Chief State School Officers annual conference on large-scale assessment, Boston.
- Crane, J. (2002, November). *The promise of value-added testing*. Washington, DC: Progressive Policy Institute. Retrieved from http://www.ppionline.org/documents/Value_Added_Testing.pdf on August 6, 2004.
- Doran, H. C. (2003, November). Adding value to accountability. *Educational leadership*, 61(3), pp. 55–59.
- Drury, D. & Doran, H. C. (2003, January). The value of value-added analysis. *National School Boards Association policy research brief* 3(1), pp. 1–4.
- Goldschmidt, P., Choi, K., & Martinez, F. (2004). *Using hierarchical growth models to monitor school performance: The effects of the model, metric and time on the validity of inferences*. Paper presented at the Council of Chief State School Officers annual conference on large-scale assessment, Boston.
- Herdman, P. A., Smith, N., & Doran, H. (2002, December). *CFNN policy brief. Value-added analysis: A critical component of determining adequate yearly progress (AYP)*. St. Paul, MN: National Charter School Alliance.
- Hershberg, T., Simon, V. A., & Lea-Kruger, B. (2004). *Measuring what matters*. National School Boards Association. Retrieved from <http://www.asbj.com/2004/02/0204asbjhershberg.pdf> on July 26, 2004.

- Kupermintz, H. (2002). *Teacher effects as a measure of teacher effectiveness: Construct validity considerations in TVAAS (Tennessee Value-Added Assessment System)*. Los Angeles, CA: National Center for Research on Evaluation, Standards, and Student Testing (CRESST).
- Kupermintz, H. (2003, Fall). Teacher effects and teacher effectiveness: A validity investigation of the Tennessee value added assessment system. *Educational evaluation and policy analysis*, 25(3) pp. 287–298.
- Marzano, R. J., Pickering, D. J., & Pollock, J. E. (2001). *Classroom instruction that works: Research-based strategies for increasing student achievement*. Alexandria, VA: Association for Supervision and Curriculum Development (ASCD).
- McCaffrey, D. F., Lockwood, J. R., Koretz, D. M., & Hamilton, L. S. (2003). *Evaluating value-added models for teacher accountability*. Santa Monica, CA: the Rand Corporation.
- McCaffrey, D. F., Lockwood, J. R., Koretz, D. M., Louis, T. A., & Hamilton, L. S. (2004, Spring). Models for value-added modeling of teacher effects. *Journal of educational and behavioral statistics*, 29(1), pp. 68–101.
- Meyer, R. H. (1996). Value-added indicators of school performance. In Hanushek, E. A. & Jorgenson, D. W. (Eds.). *Improving America's schools: The role of incentives*. Washington, DC: National Academy Press.
- Nitko, A. J. (2004). *Educational assessments of students* (4th ed.). Englewood Cliffs, NJ: Prentice Hall.
- Olson, L. (1998, May 13). A question of value. *Education week* 17(35). Retrieved from <http://www.edweek.org/ew/vol-17/35value.h17> on May 24, 2004.
- Rubin, D. B., Stuart, E. A., & Zanutto, E. L. (2004, Spring). A potential outcomes view of value-added assessment in education. *Journal of educational and behavioral statistics*, 29(1), pp. 103–116.
- Shavelson, R. J., & Towne, L. (Eds.). (2002). *Scientific research in education*. Washington, DC: National Academy Press.
- Stone, J. E. (1999). Value-added assessment: An accountability revolution. In Kanstoroom M. and Chester, E. F. Jr. (Eds.). *Better teachers, better schools*. Washington, DC: Thomas B. Fordham Foundation. Retrieved from http://education-consumers.com/articles/value_added_assessment.shtm on March 15, 2004.
- Summers, A. A. (2002, Summer). Expert measures. *Education next*. Retrieved from <http://www.educationnext.org/20022/16.html> on March 31, 2002.



Wenning, R. J., Herdman, P. A., & Smith, N. (2003, August). No Child Left Behind: Testing, reporting, accountability. *ERIC clearinghouse on urban education digest, Number 185*.

Vaughan, A. C. (2002, Spring). Standards, accountability, and the determination of success. *The educational forum, 66(3)*, p. 206- 213.

Additional copies of this and related documents are available from:
Harcourt Assessment, Inc.
19500 Bulverde Road
San Antonio, TX 78259
1-800-211-8378
1-877-576-1816 (fax)
<http://www.harcourtassessment.com>