Outputs as Educator Effectiveness in the United States: Shifting Towards Political Accountability

This manuscript has been peer-reviewed, accepted, and endorsed by the National Council of Professors of Educational Administration (NCPEA) as a significant contribution to the scholarship and practice of school administration and K-12 education.

Jody S. Piro
Texas Woman's University
Laurie Mullen
Ball State University

The definition of educator effectiveness is being redefined by econometric modeling to evidence student achievement on standardized tests. While the reasons that econometric frameworks are in vogue are many, it is clear that the strength of such models lie in the quantifiable evidence of student learning. Current accountability models frame accountability in terms of educator effectiveness through student achievement as edunomic outputs. There have been three phases of edunomic outputs, with the unit of analysis ranging from broad, institutional data to individual teachers and students. This trend for using quantitative outcomes in educational accountability is buoyed by a shifting perspective of accountability informed from those inside the profession to a definition shaped by external perspectives defined by econometric models. Educator accountability is evolving into a political accountability policy (McDonnell, 1994) as education witnesses a transition from professional to political accountability models. Educators must continue to problematize the outputs measurement for effectiveness to include broader forms of student achievement and find ways to refine measurements in econometric models for political accountability that speak to student achievement informed by professional judgment.

NCPEA International Journal of Educational Leadership Preparation, Vol. 8, No. 2– October 2013
ISSN: 2155-9635 © 2013 National Council of Professors of Educational Administration

Introduction

Accountability abounds. Professional discourse is full of references to new accountabilities, and the general citizenry served by professions are increasingly aware of and vocal about issues surrounding accountability. Accountability is the undercurrent of professional activity. The reasons that educators are immersed in increased accountability are historically situated and complex in nature. Highly public failures of trust in the business, political, and medical sectors along with the current economic crisis, contribute to our society's interest in new or redefined accountability structures.

The current narrative around accountability from the public, academics, politicians, and policymakers highlights the non-neutrality and variability of use. In its most dispassionate interpretation, accountability is construed as a simple accounting or explanation of an event – one in which there is not an expectation or culpability implied. On the other extreme, accountability is swiftly becoming synonymous with dissatisfaction, punishment, or high-stakes consequences. As Stone (1997) suggested, the accountability narrative is situated in societal and political realities with interpretations emerging from the *polis* or community at large, and interpretations will be always in flux. We contend that educator effectiveness is being redefined by an evolving accountability system focused on student performance outcomes and gauged through emerging econometric models. In this investigation of policy, we describe how educator effectiveness has been measured to suggest that the traditional model of professional accountability is shifting to a political accountability model. This shift is due, in part, to a confluence of changing social realities. Concomitantly to a shift in accountability perspectives for educators, the field has also seen an increase in the use of econometric models, both in terms of principles of the market and methodological contributions, to educational policy. While the reasons that econometric frameworks are in vogue are many, it is clear that the strength of such models lie in the quantifiable evidence produced (Dearden, Machin, & Vignoles, 2009).

Defining Teacher Effectiveness: Inputs, Processes and Outputs

In the United States, classroom teachers historically have been held accountable through school (Danielson & McGreal, 2000) or corporation teacher evaluation systems. However, the definition of teacher effectiveness has changed over the years and was typically fashioned by the dominant description of classroom effectiveness at that time (Campbell, Kyriadkides, Muijs, & Robinson, 2003; Cheng & Tsui, 1999; Cruickshank & Haefele, 1990; Muijs, 2006). Goe, Bell and Little (2008) identified the multiple measures used to evaluate teachers, including observation, principal evaluation based upon observation and other informal data, instructional artifacts, portfolios, self-reporting, student ratings, and student learning. These authors suggest that teachers' effectiveness for accountability has been measured using three distinct variables: input variables, process variables, and output or product variables. The definition of teacher effectiveness in input, process, or output models depends on distinct measures to judge effectiveness (Harris & Sass 2009; Harris & Rutledge, 2010).

The input measure of teacher effectiveness attempts to answer the effectiveness issue by isolating the variables that teachers bring to the occupation. Examples of input

measures include teacher individual characteristics, scores for college entrance examinations, high school and university GPA, achievements gained from professional teacher education (content and pedagogical knowledge), and certification or licensure status. Historically, input measurements revolved around qualities such as teacher morality and basic teaching skills (Sedlak, 1989). Teachers were expected to maintain societal mores and values as measures of their effectiveness. Later, teacher preparation and advanced degrees became the input measures used to evaluate teacher effectiveness (Darling-Hammond & Youngs, 2002; Wilson, Flooden, & Ferrini-Mundi, 2001). Teacher effectiveness research of the 1960's and 1970's promoted the clinical supervision model that attempted to demonstrate that positive teacher behaviors or personality characteristics were linked with student achievement and that certain evaluation instruments could be designed to identify those teacher behaviors (Danielson & McGreal, 2000). These presage variables, more commonly known as teacher traits, determined the effectiveness of teachers in an input driven effectiveness and accountability model (Rosenshine & Furst, 1971; Brophy & Good, 1986; Slavin, 1987; Gagne, 1977).

While input measures focused on variables solely attributed to the classroom teacher, process variables focused more on teacher and student interaction in the classroom. Termed as process-product measurement, Anderson, Evertson, and Brophy (1979) describe these processes as the "relationships between what teachers do in the classroom (the process of teaching) and what happens to their students (the products of learning)" (p. 193). Researchers identified specific teaching behaviors that had a strong correlation to higher student achievement therefore changing the definition of effectiveness to what teachers *did* rather than who they *are* (Medley, 1979). A sample of teaching behaviors that emerged from this work included clarity, variability of instruction, enthusiasm, a task-oriented approach, and others. Thus, effective teaching could be codified and taught to new and practicing teachers.

The third teacher effectiveness measure documented by Goe, Bell and Little (2008) has been an output variable. The range of output variables included student scores from teacher or district made tests, such as criterion referenced tests created by teachers; student engagement measures, such as time on task; student behaviors or social and emotional learning outcomes and student attitudes; and standardized tests such as the Stanford Achievement Test, Iowa Basic Skills, Terra Nova tests or standards-based state tests. Widespread public reporting of other standardized testing, including the National Report Card (National Assessment of Educational Progress) and international tests such as the Trends in International Math and Science Study (TIMSS) and the Program for International Student Assessment (PISA), have heightened the public's awareness of the United States' ranking in international comparisons, leading to a stronger focus on student outputs for educator accountability.

It is on this third measure of student outcomes that we focus. A shift has occurred in the field of teacher effectiveness to isolate student achievement on standardized test scores. What makes this third measure for defining teacher effectiveness different from earlier definitions of outputs is the usage of high stakes student standardized test score outputs as a primary measure for teacher accountability. According to this most recent outputs paradigm of teacher effectiveness, student achievement should be the dominant measure of teacher effectiveness rather than focusing upon specific teacher behaviors,

dispositions as measures of teacher quality, or measures of inputs, such as teacher certification (Rockoff, 2004; Rivkin, Hanushek & Kain, 2005; Harris & Sass, 2008). Lines of inquiry on program structure for traditionally or alternatively certified teachers furthered the notion of student achievement outcomes as a valid measure of teacher effectiveness (Boyd, Grossman, Lankford, Loeb, & Wyckoff, 2008; Kane, Rockoff, & Staiger, 2006). Teacher accountability, as defined by the outcomes of standardized testing, is fast becoming the dominant criterion for teacher effectiveness. We argue that policy makers have embraced powerful statistical models, influenced from the discipline of economics, that have emerged concurrently with public policy and that function as a primary accountability lever for educators.

The Influence of Economics in Defining Educator Effectiveness

Economists in the United States and Europe have studied varying dimensions of education for decades. For example, interest in the relationship between economic issues and education has driven international research aimed at understanding the impact of education in developing areas and the impact of poverty and inequalities on sustainable development (Knight, 2011; Power & MacLeans, 2010). In addition, investigations have focused upon working conditions as they intersect with teacher quality (Hanushek & Rivkin, 2007) which include: teacher pay and student achievement (Hanushek, Kain, & Rivkin, 1998); how peers affect students' achievement (Hanushek, Kain, Markman, & Rivkin, 2008); the effect of teacher absences (Clotfelter, Ladd, & Vigdor, 2009), and other areas with varying degrees of success at affecting educational accountability systems.

According to Dearden, Machin, and Vignoles (2009) an upsurge of interest in education among economists, particularly in the United Kingdom and the United States, is attributed to a trend towards regulated markets in the public domain. The authors documented the several areas where economics has impacted education. The first was in improving understanding of the impact of education on individuals and the economy as a whole such as the return on investments in education (Harmon & Walker, 1995). A second contribution was the modeling of education production and distribution of resources, such as the effects of smaller class sizes (Todd & Wolpin, 2007). The third contribution to education from economics centered on the role of education in promoting or preventing social mobility and inequality (Blanden, Gregg & MacMillan, 2008). Economists, however, have not historically studied the definition of teacher effectiveness. With the emergence of large data sets following No Child Left Behind (NCLB) and other state initiatives, such as the Tennessee Value Added Assessment System, in the United States econometric models have been used for accountability purposes for educators.

Econometric Models for Measuring Output Accountability

Econometrics is defined as an economic measurement but as the following definitions indicate, the scope is much broader. Tinter (1968) defined econometrics as "the result of a certain outlook on the role of economics, consisting of the application of mathematical statistics to economic data to lend empirical support to the models constructed by mathematical economics and to obtain numerical results" (p.74). As a social science,

econometrics is defined as a methodology "in which the tools of economic theory, mathematics, and statistical inference are applied to the analysis of economic phenomena" (Goldberg, 1964, p.1). There are several forms of econometric models utilized for quantifying student learning used in the United States. This section will address econometric models from large data sets to measure outputs to define educator effectiveness and will also focus on the concerns for using these models for measuring accountability.

With state and national reform, large data sets of student performance required quantitative models for analysis. Status models (Betebenner, 2009) create a snapshot of student growth with no comparisons made to other years of the student's learning. Improvement models (Hull, 2007) use norm-referenced measures to compare students to similar cohorts. Growth models (Gong, 2003) measure student learning as it changes from point to point in time. Unlike the achievement levels of minimum student performance mandated by No Child Left Behind (2002), growth models calculate progress or pace in order to summarize and project trends in student outputs to provide information about which programs or teachers are "growing" students (Betebrenner, 2009; Briggs & Weeks, 2009).

The focus of measurement within growth models varies, depending on the type of growth model that is used: growth-to-standard, simple growth or projection growth. A growth-to-standard model demonstrates criterion-referenced goals by measuring student progress toward proficiency of standards, usually with a cut score that identifies proficiency of student growth (Hull, 2007). In simple-growth models, norm-referenced measures compare the student to herself as she moves from grade to grade and suggest academic growth (Hull, 2007). In simple-growth models, students may demonstrate individual learning but are not compared to a criterion, such as in growth-to-standard models

In recent years, a third growth model category, projection growth, has proliferated policy-oriented educator accountability. Projection-growth models use the student's previous achievement levels from standardized tests to predict future learning levels of growth (Betebenner, 2009). They predict the change that should occur over time in the student's academic growth based upon previous performance. There are several projection-growth models, but value-added models (VAM's) have been most influential in policy aimed at educator accountability.

Value-added models are a collection of complex, statistical techniques that use multiple years of student test score data to estimate the effects of individual schools or teachers (Sanders & Rivers, 1996). Tennessee was a leader in the value-added movement with students measured for academic growth in grades three through high school using William Sanders' value added model, the Tennessee Value Added Assessment System (TVAAS) (Sanders & Rivers, 1996), which is now known as Education Value Added Assessment System (EVAAS). William Sanders, a statistician at the University of Tennessee, modified statistical models he had previously applied to agricultural genetics to "enable a multivariate, longitudinal analysis of student achievement data," (Sanders & Horn, 1998, p. 2). Sanders and his research team have claimed that given several years of student test data, they could predict a student's growth pattern (Sanders & Horn, 1998). Ohio subsequently adopted statewide value-added measures (Braun, Chudowsky, & Koenig, 2010) and value-added growth models are used in Pennsylvania and North

Carolina (Eckert & Dabrowski, 2010). Furthermore, both Dallas, TX (Thum & Bryk, 1997) and Milwaukee, WI (Meyer & Dokumaci, 2010) adopted value-added models for measuring their students' performance outputs.

Holding teachers accountable for this redefined notion of effectiveness as measured by the outcomes of students and quantified through econometric frameworks has not been without its detractors. Concerns over statistical modeling to measure educator effectiveness have emerged over recent years as researchers problematized their usage for accountability. Rand Corporation researchers, McCaffrey, Koretz, Lockwood, and Hamilton (2004), conducted extensive research using value-added models and explored the benefits and perils of their use for teacher accountability. Concerns over econometric models such as value-added growth models were varied, but included: the dilemma of missing or corrupt data, such as when a student transfers into a new school district; the difficulty of assigning teacher effect scores to one person when a team of people are associated with a students' learning, such as when teachers team teach in inclusion classrooms; the variability of class size and its relationship to student achievement; and the inconsistency of value-added scores from year to year (Amrein-Beardsley, 2008; **Baratz-Snowden, 2009;** McCaffrey, Lockwood, Koretz, Louis, & Hamilton, 2004; RAND, 2004).

While Sanders and Horn (1998) have contended that value-added models control for outside influences, such as socio-economic factors, peers, and school factors, other researchers have cautioned against using value-added models to measure student achievement for teacher evaluations based on reliability and validity arguments (Braun, 2005; Kupermintz, 2003; Lockwood, Louis & McCaffrey, 2002). Braun (2005) cautioned that the problems associated with value-added measures cannot be overstated because scores at the extreme ends of the scale are not reliable in indicating teacher quality. According to Braun (2005) to obtain proper estimates of teacher effects the ideal setting is

> …a school system in which, for each grade, students are randomly grouped into classes and teacher in that grade are randomly allocated to those classes. Roughly speaking, randomization levels the playing field for all teachers in that each teacher has an equal chance of being assigned to any class. (p. 7)

According to Hill (2000), William Sanders seemed to caution against the use of scores to evaluate individual teachers. He suggested that releasing teacher scores publically would be inappropriate because the purpose of value-added models is for school improvement, not embarrassing teachers. Regardless of Sander's caution, the Los Angeles Times (2010) released individual value added scores of classroom teachers stating that the public had the right to know the performance of public employees. To add to the mounting list of complications associated with using value added models for teacher accountability is the fact that approximately 69% of teachers cannot be accurately assessed using valued added measures (Prince, Schuermann, Guthrie, Witham, Milanowski, & Thorn, 2006). Nevertheless, despite these limitations econometric models have continued to make inroads into teacher effectiveness accountability. Researchers have maintained that while econometric modeling for the purposes of teacher effectiveness is flawed methodologically, the measures were still significant for policy making because all

measurements have inherent weaknesses (Ballou, 2005). Clearly, the growth of econometrics to investigate the impact of student academic growth has dramatically affected the notion of educator effectiveness for accountability purposes.

Recent policy changes regarding educator effectiveness went beyond linking economics and education and attempted to frame teacher and school accountability using statistical modeling with large-scale student achievement data sets as the key variable to hold individual teachers and schools accountable in an outputs version of teacher effectiveness. This definition analyzes the impact of student growth via standardized tests and traces that achievement back to schools and districts, teacher education institutions, and individual teachers to measure the effect these entities have on student academic growth and even on the long-term impact of students' earnings. The result of economic modeling on redefining the notion of teacher effectiveness is evident, and public policy resulting from statistical modeling of student achievement has been widespread. Public policy aimed at measuring educator effectiveness as defined by the outputs of student achievement via econometric modeling can be considered in three phases. Phase one is large units of analysis of edunomic outputs. Phase two is educator effects as edunomic outputs. Phase three is long-term student edunomic outcomes (see Table 1).

Output Phases for Accountability: Trend to Edunomic Outputs

Recent policy aimed at collecting and analyzing student outcomes has led to the creation of large data sets and subsequent models for analysis that are influenced by economics supporting the third definition of educator effectiveness as one that measures outputs. What is unique about the output phases is the unit of analysis employed to define educator effectiveness, units ranging from district level data to current models that narrow to focus on individual teachers and long-term student outcomes. The phases demonstrate the migration to an increasingly accepted use of econometric models to define effectiveness and accountability. We use the term *edunomic output* to refer to the definition of educator effectiveness defined by econometric models. Other fields, such as physics, coined econophysics as a new area of research developed recently between economists and physicists to understand empirical and statistical modeling of financial markets (Mantegna & Stanley, 2000; Stauffer, 2004-5; Stanley, Amarala, Gabaixb, Gopikrishnana & Plerou, 2001; Gallegati, Keen, Lux, & Ormerod, 2006). Similar to education, econophysics emerged from the availability of large-scale data sets.

The use of econometric models is shaping the definition of educator effectiveness as a new form of output measure. The first phase of edunomic outputs focused upon large sets of data at the state and the local school district level for accountability. The second phase of edunomic outputs used the quantifiable student achievement outcomes from phase one and other data bases to connect an individual teacher to the growth of learning stimulated by that teacher. The third phase used edunomic outputs that spotlight the long-term outcomes of individual students, such as earnings and educational levels, to classroom teachers' effectiveness.

Phase One: Large Units of Analysis of Edunomic Ouptuts. In the first phase, edunomic outputs defining teacher effectiveness for accountability focused on large units of analysis. This phase used quantifiable outputs from state testing initiated by The No

Child Left Behind Act (2002) that held states accountable for student achievement. Under this pioneering regulatory legislation for educational accountability, states were required to develop annual assessments in reading/language arts, math and science. For the first time, schools were mandated to publically report student achievement progress in degrees of proficiency. Additionally, school districts were obliged to report Adequate Yearly Progress (AYP) at the school, district, and state levels, including the student achievement on standardized tests for various subgroups. Large district-level data sets of student achievement in the form of standardized test outcomes were reported to the state and to consumers. Individual assessment results were available to students and parents, but the focus for accountability fixed upon aggregate school level data.

With its focus on large-scale institutional data sets as the unit of analysis, the outcome of the No Child Left Behind (NCLB) legislation represented the first phase of econometric regulation for educational accountability. Within this first phase of edunomic outputs, econometric models were introduced for measuring student achievement. Econometric models for measuring student academic success gauged AYP for NCLB (U.S. Department of Education, 2008). As policy makers and the public question teacher effectiveness, the institutions that prepare classroom teachers are also criticized for perceived deficiencies of quality. Teacher effectiveness output models measuring the impact of individual teacher education institutions on the achievement of graduates' students according to standardized tests have been developed. In effect, these new measures have equated the current model of K-12 teacher effectiveness for accountability with teacher education accountability. Student achievement outcomes, in part measured by standardized testing, are now being heralded as the measure of effectiveness in an edunomic output accountability model for both classroom teachers and the educator programs that trained them.

Both Tennessee and Louisiana currently use value added measures to rate their teacher education institutions. In these states, students' level of growth is traced back to specific teachers which is further traced back to the teachers' preparation program, resulting in a rating of teacher education institutions that is available to consumers in those states (Noell & Burns, 2006; First to the Top, 2010). The Obama administration appears to be aiming for all teacher education institutions to join Tennessee and Louisiana in reporting how their graduates affect student learning outcomes with value added measures through the Higher Education Act (Sawchuck, 2012). Delaware, Florida, Georgia, Hawaii, Maryland, Massachusetts, North Carolina, New York, Ohio, Rhode Island, Texas and the District of Columbia are reporting or plan to report value added measures for their teacher preparation programs to the public (Crowe, 2011; Sawchuck, 2012). Indeed, the U.S. Department of Education (2011) proposed student growth of elementary and secondary school students taught by program graduates as one of three outcomes-based measures of effectiveness for teacher education programs. When student growth is the measure of teacher effectiveness as measured by econometric models, teacher education institutions may also be regarded within the first phase of econometric models for accountability.

Phase Two: Educator Effects as Edunomic Outputs. While No Child Left Behind (2002) introduced district and school accountability as measured by standardized test data, the Race to the Top competition introduced the second phase of quantitative models for teacher accountability. The Race to the Top legislation defined effective teachers as

"teachers whose students achieve acceptable rates (at least one level of an academic year) of student growth" and provided substantive monies to adopt statewide policy, in part to evaluate teachers with student achievement data (U.S. Department of Education, 2009, p. 12). Since Race to the Top legislation was enacted, nearly half of the states have passed legislation allowing or mandating student achievement to become a component of teacher and principal evaluations (Piro, Wiemers, & Shutt, 2011). This trend toward regulation of teacher evaluations is gaining traction as a valid social and regulatory measurement of educational accountability. These quantitative measures have radically impacted society's views of teacher effectiveness and of the ways they should be held accountable. This second phase of outputs for accountability ties individual teacher effects to student achievement and personal teacher and principal evaluations. Race to the Top legislation advanced the use of student learning outcomes to measure teacher effectiveness by stating that "effective teachers and principals" were those who positively impacted student growth.

Reauthorizing the No Child Left Behind Act has a continued a focus on large aggregate school data but narrows the unit of analysis to teachers as well. The language in the reauthorization of the Elementary and Secondary Education Act (ESEA) states that Local Educational Agencies should "use multiple valid measures in determining performance levels, including as a significant factor data on student growth for all students" to evaluate teachers (U.S. Department of Education, 2012, p. 6). Additionally, turnabout principles for schools should control for the quality of the teachers by determining their effectiveness for instruction (p. 10). The act defines student growth as a change in student learning from time to time and regulates that high quality assessments must be used to quantify an accurate measure of that student growth over a full academic year or course (p. 8). Thirty-five states and the District of Columbia have received waivers for flexibility in requirements for NCLB (U.S. Department of Education, 2012b). One of the waiver requirements was to implement a state-wide teacher evaluation system. The second phase of teacher effectiveness edunomic outputs for accountability further narrowed the focus of analysis for teacher effectiveness outputs for measuring accountability from large aggregate institutional data to measuring individual teacher effects on student achievement via standardized tests through econometric models.

Third Phase: Long-Term Student Edunomic Outcomes. The third phase of outputs for educator accountability follows teacher effects through growth models to predict the long-term outcomes of K-12 students in output measures that highlight the longitudinal effects of teachers on students' lives. Haycock and Hanuschek (2010) stated that providing top quartile rather than bottom quartile teachers for four years would completely close the achievement gap between Caucasian and African American students. Harvard researchers used this focused unit of analysis and found that value-added teacher effects related to long term student outputs, such as higher education and the value of students' lifetime income. Specifically, they found that:

> students assigned to high-VA teachers are more likely to attend college, attend higher-ranked colleges, earn higher salaries, live in higher SES neighborhoods, and save more for retirement. They are also less likely to have children as teenagers. Teachers have large impacts in all grades from 4 to 8. On average, a one standard deviation improvement in teacher VA in a single grade raises

> earnings by about 1% at age 28. Replacing a teacher whose VA is in the bottom 5% with an average teacher would increase the present value of students' lifetime income by more than $250,000 for the average classroom in our sample. (Chetty, Friedman, & Rockoff, 2011, p. 1)

Horwitz and Ballou (2012) suggested that the Chetty, et al. conclusions may need further investigation as to the possible bias between high value added teachers and their students' long-term outcomes. As the debate continues, what is clear is that the unit of analysis for accountability has shifted in focus from large-scale units of analysis (Phase one), to individual teacher focus (Phase two), to the long term effects of a student's teachers (Phase three).

Summary of Three Phases of Edunomic Output

To summarize, in the first phase of edunomic outputs, federal and state level regulatory entities implemented a unit of analysis that focused upon data at the institutional level, such as school and district level or teacher education program level in K-12 education. In the second phase, there was a narrowing of the unit of analysis from broad institutional data to individual teacher data as regulatory bodies attempt to spotlight the role of teacher effects in student achievement outcomes. In the third phase of accountability using edunomic outputs, teacher effects are followed longitudinally through valued-added growth models to the most focused unit of analysis to date the long-term outcomes in student lives. No regulation has emerged from Phase three.

Table 1
Phases of Educator Effectiveness Edunomic Outputs for Accountability

Edunomic Output Phases	Unit of Analysis for Accountability	Sponsor	Organization Measured	Measure
First Phase	School Districts, Schools	Federal Government (No Child Left Behind and Reauthorization)	K-12	Standardized tests scores emerging from state standards
	Teacher Education Institutions	Federal Government (Higher Education Act) and States	Higher Education	Standardized test scores, traced from student to teacher to teacher education preparation program
Second Phase	Individual Teachers and Principals	Federal Government (Race to the Top; Reauthorization of No Child Left Behind)	K-12	Standardized test scores; other quality assessment data

	Individual Teachers	States (State of Tennessee; State of Ohio; State of Pennsylvania); Municipalities (Dallas, Milwaukee)	K-12	Standardized test scores of students traced to individual teachers
Third Phase	Individual Students	None to date.	K-12	Standardized test scores traced to individual teachers, then to long-term student outcomes (earnings, education, etc.)

Evolution Towards Political Accountability

Current accountability models frame accountability in terms of educator effectiveness through student achievement as edunomic outputs. This trend is buoyed by a shifting perspective of accountability informed from those inside the profession to a definition shaped by external perspectives and defined by economic and quantitative models. In what follows, we discuss the evolving model of educator accountability as it transforms from one of self-regulating autonomy to one defined by outside perspectives with the effect of regulatory policy.

McDonnell (2000) posited two models of accountability at the organizational level that serve as a framework for discussion. *The professional model* of accountability bases its claims on members' expert mastery of a specialized body of knowledge. The complexity of the field in professions requires that application of that knowledge should be regulated by a code of ethics internal to the profession and by the voluntary groups representing it. In this view, the application of professional knowledge to individual clients' needs requires judgment, so it cannot be reduced to rules or prescriptions for practice; thus, professionals require autonomy from external political control in determining how the products of their expertise should be used. Until recently, those inside the profession defined the knowledge and skills necessary for effective teaching. Advocating that teaching is both an art and science, those inside of the profession argue that while specific teaching skills can be taught, a dimension of judgment and subjectivity remain key to effective classroom instruction.

In contrast, the *political model of accountability* holds that a larger public interest transcends the interests and values of any single class or person, and that its pursuit is best ensured if individual behavior is held accountable to the larger body politic. Dissatisfaction with the dominant professional model of accountability reflects the political model of accountability in education in the United States. Consequently, it argues not only that public employees' behavior should be constrained, but that the constraints should be externally imposed. The emphasis on edunomic outputs to define school and educator effectiveness is evidence of the transition to a political accountability model. Entities outside of the profession, in this case the federal and state government, imposed standardized testing as the ultimate accountability measure. As a result, effective teaching is now reduced to a value added score based on student testing. One of

the results of an externally imposed definition of educator effectiveness is the shaping and constraining of practice to produce higher test scores. Outputs on standardized testing are increasingly seen as evidence of effectiveness. Educators did not seek the use of value-added measures to determine effectiveness, and recent regulatory policy efforts, such as No Child Left Behind and Race to the Top, emerged from the influence of accountabilities *outside* of the profession.

McDonnell's bifurcated professional versus political models provide guidance in defining the issue. Dualities, however, rarely lead to mutually agreeable solutions. Therefore, how can an integrated model address the new political accountability as well as recognize professional expertise? Professional accountability in education need not exist separately from regulatory political accountabilities. A more productive option between accountability systems is to value and influence each other in a reciprocal system where input from one system recharges and modifies the other, as in the case of open systems (Katz & Kahn, 1978). Currently, professional and political accountability models, built with the purpose of persuading people and gaining (or maintaining) power, are either in conflict or coexist without acknowledging the other system. In the language of systems theory, they are closed systems (Katz & Kahn, 1978). There is little exchange between organizations, and ideas about accountability are separated from the overall environment.

Professional accountability in education, often influenced by progressivism and student-oriented pedagogies, maintained the inward focus towards professional accountability. While outside influences mandated reform resulting in a new form of accountability, classroom teachers and administrators can provide valuable expertise and specialized knowledge to contribute to the complexity of defining and maintaining educational excellence. To do so, educators must seek ways to transform these valid dimensions of professional practice into models that have similar policy-leveraging effects and that mutually serve the political accountability model. Student achievement should remain at the forefront of professional accountability. Expanded metrics to define student achievement from its current interpretation will capture additional values that are important to the profession and society. For example, in the reauthorization of the Elementary and Secondary Education Act (ESEA) language is proposed that school corporations use multiple measures to determine student achievement in teacher evaluation. Additionally, the Bill and Melinda Gates Foundation found that using three measures of outputs for teacher evaluation (value-added scores, observations and student feedback) provided a more balanced snapshot of how teachers would affect student achievement (Sawchuck, 2013). Involving classroom teachers and administrators in the creation, validation, and implementation of measures that contribute to the definition of student achievement and thus effectiveness is one way to recognize educator expertise as well as to support a broader definition of student success. Educators must continue to problematize the outputs measurement for effectiveness to include broader forms of student achievement and find ways to refine measurements in econometric models for political accountability that speak to student achievement informed by professional judgment.

Summary

Accountability is becoming synonymous with high-stakes consequences for teachers. The definition of educator effectiveness and the measures used to gauge that effectiveness have changed over the years and have encompassed input variables, process variables, and output variables. Recently, policy makers have embraced student testing as a primary metric for effectiveness, and as a result, edunomic outputs have become a lever for accountability purposes. Powerful statistical models influenced from the discipline of economics have been developed and evolved concurrently with public policy aimed at accountability for educators. These models fall into three phases with increasingly narrow units of analysis. In the first phase of edunomic outputs, federal and state level regulatory entities implemented a unit of analysis that focused on data at the institutional level. In the second phase, there is a narrowing of the unit of analysis from broad institutional data to individual teacher data as regulatory bodies attempt to spotlight the role of teacher effects in student achievement outcomes. In the third phase of accountability using edunomic outputs, teacher effects are followed longitudinally through valued added growth models to the most focused unit of analysis to date in the outputs measure of teacher effectiveness – the long-term outcomes in a student's life.

In a relatively short period of time, the professional model of accountability for educators has transformed from one of self-regulating autonomy to a political model defined by outside perspectives (McDonnell, 1994). Combined with increased accountability expectations for all professions in a time of economic turmoil, the allure of the possibilities offered by the various growth models is changing the education landscape. The future reach of econometric models to define educator effectiveness is unknown, but as researchers continue to refine and strengthen these methodologies, we can be sure of edunomic outputs as a significant dimension of educator effectiveness for many years to come. At first glance, it appears that traditional professional accountability models will be replaced by a political model led by perspectives defined by those outside of the profession, but to continue viability of a professional accountability, it is inherent for educators to discover the ways that incorporate and influence accountability measures. Educators must embrace such measures as a key dimension of professional accountability to inform design, delivery, and evaluation of programs that prepare students for the 21st century.

References

American Association of Colleges for Teacher Education. (2011). *Transformations in educator preparation: Effectiveness and accountability*. Washington, D.C, Author.

Amrein-Beardsley, A. (2008). Methodological concerns about the education value-added assessment system. *Educational Researcher*, *37*, 65-75. Retrieved from http://dx.doi.org/10.3102/0013189X08316420

Anderson, L., Evertson, C. & Brophy, J. (1979). An experimental study of effective teaching in first-grade reading groups. *Elementary School Journal, 79*(4), 193-223.

Ballou, D. (2005). Value-added assessment: lessons from Tennessee. In R.W. Lissitz (Ed.), *Value-added models in education: Theory and applications* (pp. 1–26). Maple Grove, MN: JAM Press.

Baratz-Snowden, J. (2009). *Fixing tenure: A proposal for assuring teacher effectiveness and Due process.* Washington, DC: Center for American Progress.

Betebenner, D. (2009). Norm-and criterion-student growth. *Education Measurement: Issues and Practices*, *28*(4), 42-51.

Blanden, J., Gregg, P. & MacMillan, L. (2008). *Intergenerational persistence in income and social class: The impact of increased inequality*. CMPO mimeo.

Boyd, D., Grossman, P., Lankford, H., Loeb, S., & Wyckoff, J. (2008). *Teacher preparation and student achievement* (NBER Working Paper No. W14314). Cambridge, MA: National Bureau of Economic Research.

Braun, H.I. (2005). *Using student progress to evaluate teachers: A primer on value-added models*. Princeton, NJ: Educational Testing Service.

Braun, H., Chudowsky, N., & Koenig, J. (2010). *Getting value out of value-added Report of a workshop*, Retrieved from http://216.78.200.159/Documents/RandD/Other/Getting%20Value%20out%20of%20 Value-Added.pdf

Briggs, D., & Weeks, J. (2009). The impact of vertical scaling decisions on growth interpretations. *Educational Measurement: Issues and Practice*, *28(4),* 3-14.

Brophy, J., & Good, T. (l986). Teacher-effects results. In M.C. Wittrock (Ed.), *Handbook of research on teaching.* (3rd ed., pp. 328-376). New York: Macmillan.

Campbell, R. J., Kyriakides, L., Muijs, R. D., & Robinson, W. (2003). Differential teacher effectiveness: Towards a model for research and teacher appraisal. *Oxford Review of Education*, *29*(3), 347-362.

Cheng, Y. C., & Tsui, K. T. (1999). Multi-models of teacher effectiveness: Implications for research. *The Journal of Educational Research, 92*(3), 141-150.

Chetty, R., Friedman, J., & Rockoff, J. (2011). *The long-term impacts of teachers: Teacher value-added and student outcomes in adulthood* (Working Paper). Harvard University. Retrieved from http://www0.gsb.columbia.edu/faculty/jrockoff/papers/value_added_dec2011.pdf

Clotfelter, C. T., Ladd, H. F., & Vigdor, J. L. (2009). Are teacher absences worth worrying about in the United States? *Education Finance and Policy*, *4*(2), 115–149.

Cochran-Smith, M., & Fries, M. (2001). Sticks, stones, and ideology: The discourse of reform in teacher education. *Educational Researcher, 30*(8), 3-16.

Cochran-Smith, M. (2004). Defining the outcomes of teacher education: What's social justice got to do with it? *Asia-Pacific Journal of Teacher Education, 32*(3), 193–212.

Crowe, E. (2011). *Race to the top and teacher preparation: Analyzing state strategies and fostering program innovation.* Retrieved from http://www.americanprogress.org/issues/2011/03/pdf/teacher_preparation.pdf

Cruickshank, D. R., & Haefele, D. L. (1990). Research-based indicators: Is the glass half-full or half-empty? *Journal of Personnel Evaluation in Education, 4*(1), 33-39.

Danielson, C., & McGreal, T. L. (2000). *Teacher evaluation to enhance professional practice*. Alexandria, VA: Association for Supervision and Curriculum Development.

Darling-Hammond, L. (1989). Accountability for professional practice. *Teachers College Record, 91*(1), 59-80.

Darling-Hammond, L., Wise, A., Klein, S. (1999). *A license to teach: Raising standards for teaching*. San Francisco, CA: Jossey-Bass.

Darling-Hammond, L. & Youngs, P. (2002). Defining "highly qualified teachers": What does "scientifically-based research" actually tell us? *Educational Researcher*, *31*(9), 13-25.

Darling Hammond, L. (2011). Effective teaching as a civil right: How building instructional capacity can help close the achievement gap*. Voices in Urban Education, 13(*44-58*).*

Dearden, L., Machin, S., & Vignoles, A. (2009). Economics of education research: A review and future prospects. *Oxford Review of Education, 35*(5), 617-632.

Eby, M. (2000) The challenges of being accountable. In A. Brechin, H. Brown & M. Eby (Eds.), *Critical Practice in Health and Social Care* (pp. 187-208). London: Sage.

Eckert, J., & Dabrowski, J. (2010). *Should value added measures be used for performance pay?* Phi Delta Kappan. Retrieved from http://projects.brevardschools.org/PAS/Shared%20Documents/Shouldvalueadded measures.pdf

Ewing, J. (2011). Mathematical intimidation: Driven by the data. *Notices of the American Mathematical Society*, *58 (*5*),*667-673.

Farkas, S., & Duffet, A. (2010). *Cracks in the ivory tower? The views of education professor circa 2010*. Thomas P. Fordham Institute. Retrieved from http://www.edexcellencemedia.net/publications/2010/201009_cracksintheivorytower/Cracks%20In%20The%20Ivory%20Tower%20-%20Sept%202010.pdf

First to the Top Act. (2010). Pub.ch. No. 2, Senate Bill No. 7005.

Fraser, J. W. (2001). Time to cut the link between teacher preparation and certification? *Education Week*, *20*(20), 56.

Gagne, R. (1977). *Conditions of learning*. (3rd ed.). New York, NY: Holt, Rinchart, & Winston.

Gallegati, M., Keen, S., Lux,T., & Ormerod, P. (2006). Worrying trends in econophysics, *Physica A*, *370(*1*),* 1–6.

Goe, L., Bell, C., & Little, O. (2008). *Approaches to evaluating teacher effectiveness: A research synthesis.* Washington, D. C.: National Comprehensive Center for Teacher Quality.

Goldberg, A. (1964). *Econometric Theory*. New York, John Wiley & Sons.

Gong, B. (2003). Models for using student growth measures in school accountability. Center for Assessment at the Council of Chief State School Officers. Washington, D.C. Retrieved from http://www.nciea.org/publications/GongGrowthModels111504.pdf

Haunshek, E., Kain, J., & Rivkin, S. (1998). Do higher salaries buy better teachers? (mimeo). University of Rochester.

Hanushek, E., Kain, J., Markman, J., & Rivkin, S. (2008). Does peer ability affect student achievement? *Journal of Applied Econometrics, 18*(5), 527-544.

Hanushek, A., &. Rivkin, S. (2007). Pay, working conditions and teacher quality. *The Future of Children, Excellence in the Classroom*, *17*(1), 69-86.

Harris, D., & Sass, T. (2008). *Teacher training, teacher quality and student achievement* (Working Paper 3). Washington, DC: Center for the Analysis of Longitudinal Data in Education Research, Urban Institute.

Harris, D., & Sass, T. (2009). *What makes for a good teacher and who can tell?* National Center for the Analysis of Longitudinal Data in Education Research (CALDER).

Harris, D., & Rutledge, S. (2010). *Models and predictors of teacher effectiveness: A comparison of research about teaching and other occupations* (Working Paper 30). Washington, DC: Urban Institute.

Haycock, K., & Hanuschek, E. (2010). An effective teacher in every classroom. *Education Next*, *10*(3). Retrieved from http://educationnext.org/an-effective-teacher-in-every-classroom/

Hess, F. (2001). *Tear down the wall: The case for radical overhaul of teacher licensure.* Washington, D.C.: Progressive Policy Institute. Retrieved from http://www.ppionline.org/ppi_ci.cfm?knlgAreaID=110&subsecID=135&contentID=3964

Hill, D. (2000). He's got your number. Teacher Magazine, 11(8), 42-47.

Horwitz, J. & Ballou, D. (2012). Looking again at long-term teacher effects and 'value-added.' Retrieved from http://nepc.colorado.edu/newsletter/2012/02/review-long-term-impacts

House, E. R. (1993). *Professional evaluation.* Newbury Park, CA: Sage.

Hull, J. (2007). *Growth models: A guide to informed decision making*. The Center for Public
Education. Retrieved from http://www.centerforpubliceducation.org/Main-Menu/Policies/Measuring-student-growth-At-a-glance/Measuring-student-growth-A-guide-to-informed-decision- making.html

Kane, T. J., Rockoff, J. E., & Staiger, D. (2006). *What does certification tell us about teacher effectiveness: Evidence from New York City* (NBER Working Paper Series. Vol. w12155). Retrieved from http://www.nber.org/papers/w12155.

Katz, D., & Kahn, R. (1978). *The social psychology of organizations*. (2nd ed.). New York: Wiley.

Knight, J. (2011). Education hubs: A fad, brand, an innovation? Journal of Studies in International Education, 15(3), 221-240.

Kupermintz, H. (2003). Teacher effects and teacher effectiveness: A validity investigation of the Tennessee value-added system. *Educational Evaluation and Policy Analysis*, *25*(3), 287-298.

Lockwood, J., Louis, T., & McCaffrey, D. (2002). Uncertainty in rank estimation: Implications for value-added modeling accountability systems. *Journal of Educational and Behavioral Statistics*, *27*(3), 255-270.

Los Angeles Times. (2010).Grading the teachers: Value added analysis. Retrieved from http://www.latimes.com/news/local/teachers-investigation/

Mantegna R., & Stanley, H. (2000). *An introduction to econophysics: Correlations and complexity in finance.* Cambridge, MA: Cambridge University Press

McCaffrey, D. F., Lockwood, J., Koretz, D., Louis, T. A., & Hamilton, L. (2004). Models for value-added modeling of teacher effects. *Journal of Educational and Behavioral Statistics, 29*(1), 67–101.

McDonnell, L. M. (2000). Defining democratic purposes. In L. M. McDonnell, P. Timpane, and R. Benjamin (Eds.), *Rediscovering the democratic purposes of education* (pp.1-18). Lawrence: University Press of Kansas.

McDonnell, L. M. (2004). *Politics, persuasion, and educational testing*. Cambridge: Harvard University Press.

Medley, D. M. (1979). The effectiveness of teachers. In P.L. Peterson and H. J. Walberg (Eds.), *Research on teaching: Concepts, findings, and implications.* Berkeley, CA: McCutchan.

Meyer, R., & Dokumaci, E. (2010). Value-added models and the next generation of assessments. Center for K-12 Assessment & Performance Management. Retrieved from http://www.k12center.org/rsc/pdf/MeyerDokumaciPresenterSession4.pdf

Muijs, D. (2006). Measuring teacher effectiveness: Some methodological reflections. *Educational Research & Evaluation, 12*(1), 53-74.

National Commission on Teaching and America's Future. (1996). *What matters most: Teaching for America's future*. New York.

National Council for the Accreditation of Teacher Educators (2010). *Transforming teacher education through clinical practice: A national strategy to prepare effective teachers*. Retrieved from http://www.ncate.org/LinkClick.aspx?fileticket=zzeiB1OoqPk%3D&tabid=715

National Council on Teacher Quality (2011). State of the states. Trends and early lessons on teacher evaluation and effectiveness policies. Retrieved from http://www.edweek.org/media/nctq_stateofthestates_embargoed.pdf

No Child Left Behind. (2002). Public Law No. 107-110, 115 Stat. 1425.

Noell, G., & Burns, J. (2006). Value added assessment of teacher education. *Journal of Teacher Education, 57,* 37-50.

Piro, J., Wiemers, R. & Shutt, T. (2011). Using student achievement data in teacher and principal evaluations: A policy study. *International Journal of Educational Leadership Preparation, 6(4)*. Retrieved from the Connexions Web site: http://cnx.org/content/m41125/1.2/

Prince, C., Schuermann, P., Guthrie, J., Witham, P., Milanowski, A., & Thorn, C. (2006). The other 69 percent: Fairly rewarding the performance of teachers of non-tested subjects and grades. Washington, DC:, Office of Elementary and Secondary Education.

Power, C., & Maclean, R. (2010). Lifelong learning for poverty alleviation and sustainable development. *International Symposium on Lifelong Learning*, Hong Kong Institute for Education. Retrieved from http://www.ied.edu.hk/isll/Download%20Files/Panel2_Prof_Colin_Power.pdf

RAND Research Brief (2004). The promise and peril of using value-added modeling to measure teacher effectiveness. Retrieved from http://www.rand.org/pubs/research_briefs/RB9050/RAND_RB9050.pdf

Reuznel. D. (2002) Torturous routes. *Education Next, 2*(2), 42-48.

Rivkin, S., Hanushek, E., & Kain, J. (2005). Teachers, schools and academic achievement. *Econometrics*, *73*(2), 417-58.

Rockoff, J.E. (2004). The impact of individual teachers on student achievement: Evidence from panel data. *American Economic Review, 94*(2), 247-252.

Rosenshine, B., & Furst, N. (1971). Research on teacher performance criteria. In B.O. Smith (Ed.), *Research in teacher education: A symposium.* (pp. 37-72). Englewood Cliffs, NJ: Prentice-Hall.

Sanders, W., & Horn, P. (1994). The Tennessee value-added assessment system (TVASS): Mixed methodology in educational assessment. *Journal of Personnel Evaluation in Education*, *8*(1), 299-311.

Sanders, W.L., & Rivers, J.C. (1996). *Cumulative and residual effects of teachers on future student academic achievement.* (Research Progress Report). Knoxville: University of Tennessee Value-Added Research and Assessment Center.

Sanders, W., & Horn, S. (1998). Research findings from the Tennessee value-added assessment system (TVAAS) database: Implications for educational evaluation and research. *Journal of Personnel Evaluation in Education, 12*(3), 247-256.

Sawchuck, S. (2012). Value added proves beneficial to teacher colleges. *Education Week.* Retrieved from http://www.edweek.org/ew/articles/2012/02/17/21louisiana_ep.h31.html?tkn=YYOFNQPexcLYsu2fCcerGfNbY%2BmlRcgi783j&intc=es

Sawchuck, S. (2013). *Combined measures better at gauging teacher effectiveness, study finds*. Retrieved from http://www.edweek.org/ew/articles/2013/01/08/17teach_ep.h32.html?tkn=XQSF%2FRt5Y0QfRrL10tg5JOZyRSdR2Eozgzl6&cmp=ENL-EU-NEWS2

Sedlak, D. (1989). Let us go and buy a school master: Historical perspectives on the hiring of teachers in the United States, 1750 – 1980. In D. Warren (Ed.), *American teachers: Histories of a profession at work* (pp. 257-290). New York : Macmillan.

Slavin, R.E. (1987). Ability grouping and student achievement. *Review of Educational Research*, *57*(3), 293-336.

Stanley, H., Amaral, L., Buldyrev, S., Gopikrishnan, P., Plerou, V., & Salinger, M. (2002). Self-organized complexity in economics and finance. *PNAS. (USA), 99*, 2561-2565.

Stauffer, D. (2004-05). Introduction to statistical physics outside physics. *Physica A, 336*(1-2), 1-5. doi:10.1016/j.physa.2004.01.004

Stoddard, T., & Floden, R. (1995). Traditional and alternative routes to teacher certification: Issues, assumption, and misconceptions. (Issues paper 95-2). National Center for Research on Teacher Learning, Michigan State University. (ERIC Reproduction Service No. ED383697).

Stone, D. H. (1997). Policy *paradox: The art of political decision making (2nd ed.).* New York: Norton.

Tennessee Department of Education. (2007). *Tennessee's most effective teachers: Are they assigned to the schools that need them most?* (Research Brief). Nashville, TN: Author. Retrieved from http://tennessee.gov/education/nclb/doc/TeacherEffectiveness2007_03.pdf

Thum, Y. & Byrk, A. (1997). Value-added productivity indicators: The Dallas System. In J. Millman, (Ed.), *Grading teachers, grading schools: Is student achievement a valid measure?* (pp. 100-109).Thousand Oaks, CA: Corwin Press.

Tinter, G. (1968). *Methodology of mathematical economics and econometrics* Chicago, IL: University of Chicago Press.

United States Department of Education (2008). *Growth models: Ensuring grade-level proficiency for all students by 2014.* Washington, D.C. Retrieved from http://www2.ed.gov/admins/lead/account/growthmodel/proficiency.html

United States Department of Education. (2011). *Our future, our teachers: The Obama's administration plan for teacher education reform and improvement.* Retrieved from http://www.ed.gov/teaching/our-future-our-teachers

United States Department of Education. (2009). *U.S. department of education opens race to the top competition*. Retrieved from http://www.ed.gov/news/pressreleases/2009/11/11122009.html

United States Department of Education (2012). *ESEA flexibility request for window 3*. Retrieved from http://www.ed.gov/sites/default/files/esea-flexibility-request-w3-acc.doc

United States Department of Education (2012b). *Obama administration approves Idaho's request for NCLB flexibility.* Retrieved from http://www.ed.gov/news/press-releases/obama-administration-approves-idahos-request-nclb-flexibility

Wilson, S., Flooden, R., & Ferrini-Mundi, J. (2001). *Teacher preparation research: Current knowledge, gaps and recommendations*. (Document R-01-3). University of Washington: Center for Teaching and Policy.