Sean Nicholson-Crotty
University of Missouri–Columbia
Nick A. Theobald
University of Kansas
Jill Nicholson-Crotty
University of Missouri–Columbia

Organizational Performance

Disparate Measures: Public Managers and Performance-Measurement Strategies

Sean Nicholson-Crotty is an assistant professor in the Department of Political Science and a faculty research fellow in the Institute for Public Policy at the University of Missouri–Columbia. His primary research interests include public management, intergovernmental relations, and bureaucratic politics.
E-mail: nicholsoncrottys@missouri.edu.

Nick A. Theobald is a lecturer at the University of Kansas. His research includes work on state budgeting decisions, education finance reform, representation, and organizational performance.
E-mail: nickt@ku.edu.

Jill Nicholson-Crotty is an assistant professor at the Harry S. Truman School of Public Affairs at the University of Missouri–Columbia. Her research focuses on the management of public and nonprofit organizations and the interaction of these sectors in the implementation of public policy.
E-mail: nicholsoncrottyj@missouri.edu.

Over the past two decades, many studies have investigated the scope and significance of performance measurement in public organizations. Nonetheless, there is more to learn about the challenges facing public managers who want to measure organizational outputs and use the feedback to improve performance. Specifically, managers are often faced with redundant measures of the same output, each of which may be preferred by a different political principal or stakeholder. Furthermore, a manager's choice of measures can have serious consequences for both the estimation of agency problems and the success of programmatic solutions. We test these assertions in an analysis of educational organizations in Texas. We find that managers' assessments of organizational performance and decisions regarding solutions depend on the choice of performance measures.

Performance measurement has increased markedly in public organizations and has generated growing interest among scholars over the past several decades. Authors have traced the history of performance measurement (Bouckaert 1992) and articulated the value of tracking organizational performance on certain indicators (Newcomer 1997; Wholley 1999). They have also investigated the many obstacles that public organizations face when they try to develop and strategically use performance measures (Ammons 1992). Some authors have focused on the ways in which entrepreneurial public managers have taken the lead in performance measurement, whereas others have highlighted the many organizations and functions that still fail to benefit from this growing trend (Ammons 1995; Murphy 1999).

One consistent theme within this literature is the assertion of authors and practitioners alike that no single measure is sufficient to answer the many questions that good managers ask about the performance of their organizations (e.g., Kravchuk and Schack 1996). Implicit in many of these studies, however, is the assumption that a particular performance measure or set of measures can be developed to answer each *individual* question, assuming that the objective of

measurement can be clearly articulated (Kravchuk and Schack 1996; Osborne and Plastrik 2000).[1] To this end, scholars have focused on the specific purposes to which performance measures can be put (Behn 2003; Hatry 1999) or the different types of measures, including output, outcome, quality, workload, and others, that can be used to gather information about different components of public-service delivery (Berman and Wang 2000; de Lancer Julnes and Holzer 2001). The implication is that managers can select the appropriate performance measure by narrowly defining the function they want to know more about and the purpose to which they want to put that information.

We argue in this essay that the focus on different measures and different purposes, although useful, may have given us an overly simplistic view of the challenges that managers face when using performance measurement. We suggest that managers often face *multiple* plausible measurements of the same concept. In other words, it is not simply a matter of deciding to evaluate organizational performance on a particular dimension with an indicator of program outcomes and then selecting one or more measures depending on program complexity. Instead, managers may find there are three different *ways* to measure the outcome in question, even for a simple outcome such as the number of clients who drop out of a program. More important, managers may find that each measurement technique offers very different feedback regarding performance. Thus, the choices managers make about measurement can have a significant impact on their evaluations, including assessments of (1) whether their organization has a problem, (2) the environmental and organizational causes of the problem, and (3) whether their solutions to the problem are working.

The primary purpose of this essay is not to offer prescriptions concerning the best method for selecting from multiple measurement alternatives. Instead, we draw on a case from public education to illustrate the

problem of multiple measures and to explore the different types of performance feedback that managers may receive from each.

Purposes, Types, and Selection of Performance Measures in the Literature

Recent work on performance measurement has moved beyond early assessments of the prevalence and benefits of performance measurement to investigate the many purposes that these measures can be used for and the different types of measures that managers may choose from. This work is both taxonomic in nature, attempting to provide more nuance to our classifications of performance measurements, and prescriptive, offering advice on how to choose the right measure.

A number of authors have reified the myriad ways that performance measures might be used into tractable lists of purposes. Kopczynski and Lombardo (1999) list five such purposes: recognition of good performance, identification of performance targets, jurisdictional performance comparisons, accountability, and coalition and trust building. Hatry (1999) expands the list to 10 by adding internal and external budgeting activities, problem identification, evaluation, strategic planning, and improvement. Finally, Behn (2003) reduces the list back to eight by combining certain purposes under one heading. He suggests that performance measurement can be used to evaluate, control subordinates, make budgetary decisions and requests, motivate employees, promote the organization to stakeholders and political principals, celebrate accomplishments, learn about program efficacy, and improve performance.

Scholars have also sought to illuminate the different types of measures that managers might use for these varied purposes. In their analysis of performance measurement in U.S. counties, Berman and Wang (2000) identify two general categories of performance measures: those concerned with outcome or quality of service and those that measure workload factors. They suggest that the adoption and implementation of both types of measures indicate a greater depth of performance measurement use. In another study, de Lancer Julnes and Holzer (2001) further refine the types of measures into three categories. They divide performance indicators into measures of efficiency assessing costs, benefits, and outcome measures; programmatic impact and output indicators; and the policy and process components of the organization.

Finally, scholars have offered some prescriptions to managers concerning the methods by which they should develop or select measures of performance. Behn (2003) argues that the purpose to which a measure is to be put should determine its selection. For example, he suggests that an evaluation of perfor-

mance requires outcome measures, whereas budgeting and allocation decisions require efficiency measures, and so on. Kravchuk and Schack (1996) offer more specific advice to managers and organizations: They also suggest the users or purposes of the measures should be taken into account, but they go on to argue that these users should be involved in measure development. Additionally, for performance-management systems to remain relevant, managers must adapt to changing environmental and organizational conditions when revisiting old measurement choices. Finally, the authors caution against excessive aggregation of information and suggest that multiple detailed measures of the same concept are often preferable to one aggregate measure.

Managers and Multiple Measures of Performance

Although studies identifying the purposes, types, and selection of performance measures are both instructive and useful, they understate both the challenges managers face when selecting an indicator of performance and the consequences of the choices they make. Even when managers identify a specific purpose—let's say program evaluation—and settle on a particular type of indicator—an outcome measure, for instance—they are often faced with multiple ways of measuring the concept of interest. Kravchuk and Schack (1996, 357) suggest that managers must settle on an "explicit measurement strategy," but they do not explore how difficult it may be to follow that advice when numerous valid strategies exist. Perhaps the most famous example of this measurement quandary is in the area of job-training programs. If a manager decides that he or she wants to assess job-placement success, is it appropriate to measure the (1) proportion of trainees placed, (2) the stability of the employment of those who found jobs, or (3) the earnings gains of those who participated in the program? Scholars have demonstrated that the choice of measures can produce significant gamesmanship among managers and caseworkers and affects the utility of information gathered for improving organizational effectiveness (Blau 1955; Heinrich 2002).

The relative ambiguity of goals in public organizations, along with the nascent nature of performance-measurement efforts, means that multiple measures of or methods for measuring the same concept may be identified as appropriate or correct. Beyond job training, we can find examples of this in other areas of public service provision. For instance, there are multiple agreed-upon ways to measure air quality and the success of local governments in achieving clean-air standards. Similarly, economists have identified three measures of efficiency that local governments might employ in their planning and regulatory functions. These include technical and allocative efficiency, which measures the relationship between inputs and outputs within a given cycle, and dynamic efficiency, which

measures a very different concept, scarce resource usage over time (Worthington and Dollery 2000).

Finally, we can draw a couple of examples from the contentious debate over performance measurement in public education. When educational administrators or school board members want to assess whether Head Start programs within their district are doing a good job in closing the gap between black and white students, for instance, should they use the Peabody Picture Vocabulary Test or the Wechsler Preschool and Primary Scale of Intelligence? Both are widely recognized as appropriate measures, but the two may produce different findings regarding the abilities of students (Phillips, Crouse, and Ralph 1998). Alternatively, if a superintendent wants to learn about or promote the level of teacher expertise within his or her organization—Behn's seventh and fifth purposes, respectively—the literature suggests that he or she could measure the teaching staff's collective (1) subject skill, (2) education level, (3) longevity, or (4) leadership capabilities (see Brandon and Heck 1998 for a review). Each of these measures, however, will likely tell a different story about workforce expertise.

Different measures of—or different methods for measuring—concepts have real consequences for public programs if they lead managers to divergent conclusions about organizational performance. The logic of performance measurement and management suggests this may often be the case. Drucker argues that after managers define measures of performance, they "must use these measures to feed back on their efforts … to build self-control from results into their system" (1973, 158). Though scholars have warned that performance *measurement* is not always the same as performance *management*, they clearly maintain that the definition of measures has a tangible impact on the relationship between the two activities and, ultimately, on the success of the latter (Hatry 1999).

To clarify our thinking about the relationship between measurement and management, we identified three managerial decisions that might be affected by information coming from different performance measures. Obviously, these examples should be thought of as a small sample of the many things that managers do, chosen for their illustrative value. First, one of the most important management functions is the ability to identify organizational problems quickly and accurately (Altshuler 1988; Cohen and Eimicke 1995). Next, once a manager has decided that the organization has a problem, he or she must determine the causes in order to develop an effective strategy (Thompson 1967). Finally, the manager must evaluate whether the solution to the problem is working and having the desired effect on organizational outcomes (Drucker 1973; Osborne and Gaebler 1993).

The last of these tasks is the function we most often associate with performance measurement, but all three depend on accurate information about organizational performance. Unless problems are brought to the manager's attention by political principals or outside stakeholders, the manager needs performance information to identify suboptimal organizational functions. Similarly, managers must assess how the inputs into the organization, as well as the larger environment within which it works, contribute to or provide the context for the problems that they encounter. Thus, because each of these functions depends to a certain degree on performance information, we argue that the method by which important concepts are measured has a real impact on the decisions that managers make about public organizations and programs.

The Case of Public Education Organizations and Dropouts

In many ways, educational organizations are the perfect place to study the complexities of performance measurement and management. School districts are the most common type of public bureaucracy and, more important, are currently the focus of strident calls for greater organizational accountability based on outcomes.[2] In fact, education was the target of performance-based reform movements long before other public services and programs came under similar scrutiny (Ladd 1996). Performance measurement in educational organizations also raises many of the questions and controversies that mark performance-based reforms in other service areas. These include questions about which level of the organization it is most appropriate to study (Bishop 1994), whether accountability systems are sustainable or produce suboptimization on unmeasured goals (Milgrom and Roberts 1992), and how to design effective performance-measurement systems (Ladd 1996).

Most important for our purposes, education is a good place to study performance measurement because there is so little consensus on the appropriate way to measure outcomes of interest. In a review of accountability studies across multiple school districts, one notable scholar remarked that "the measurement of performance is a central feature of all of these systems. It is intriguing, therefore, that the systems fail to use a common approach for measuring school performance" (Meyer 1996). He is referring to the fact that states vary dramatically in the method by which they measure "student achievement." There are also disagreements among scholars and practitioners as to the appropriate way to measure the cost of educating each child (Duncomb, Ruggiero, and Yinger 1996), as well as the best way to compare inputs to outputs in educational organizations (Barnett 1996). Thus, educational organizations present a setting in which managers are often confronted with multiple methods for measuring the

same indicator of performance. The foregoing argument suggests the choices that superintendents make among these measures may have real consequences for the decisions they make about their organizations.

Measuring and Managing the Dropout Problem in Texas School Districts

Dropouts are one of the most salient and difficult problems for educational administrators in this country. Questions about why kids drop out, how to keep them in school, and the consequences of low educational attainment have produced a massive amount of scholarly and practitioner-oriented literature. Studies have examined the environmental causes of the decision to drop out of school, focusing on the effects of poverty, neighborhood, and social networks (Haveman and Wolfe 1995; Velez 1989). Others have focused on the organizational correlates of high dropout rates, looking for the influence of high-stakes testing, discipline policies, district wealth, and teacher characteristics on the decision to leave school (Ekstrom 1986; Fletcher 2002). Finally, a large number of studies have investigated the success of programs implemented by districts to reduce the number of students who drop out (see Orr 1987 for a review).

Obviously, dropouts are an important and widely recognized problem for educational organizations, and therefore we would expect the managers of these organizations to be keenly interested in measuring and managing performance in this area. This study will focus on the challenges that managers of school districts face in doing this in one state—Texas. The focus on a single state is necessary if we are to gain a sufficiently nuanced understanding of the specific measures available to managers, as well as a necessary appreciation for the broader context of performance measurement in these organizations. In this research design, the key characteristics of interest vary, but organizational type and structure, as well as the macro political context, are held constant. This, in fact, is one of the advantages of studying multiple organizations within a single state and provides a solid foundation for inference to organizations outside the sample (Nicholson-Crotty and Meier 2003).

Texas is an appropriate choice for our single-state study because of the tremendous variability of the educational organizations within the state. The state of Texas operates a relatively decentralized system, with most authority residing in the local school districts.[3] Each district determines its own curriculum and makes all of its own personnel decisions. The sample of districts includes great diversity; districts range from monoracial to multiracial, great affluence to considerable poverty, rural to suburban to urban, and tiny to very large. Approximately one of every 14 school districts in the United States is located in Texas. This sample provides a good base on which

to build inferences about the public management of education in the United States and the beginnings of possible inferences about public management more broadly.

Performance measurement in many public organizations is the result of legislative mandate (de Lancer Julnes and Holzer 2001), and the measurement of dropouts in Texas school districts is no exception. In 1984, the legislature passed House Bill (HB) 72, which officially defined a dropout as "a student in Grades 7–12 who does not hold a high school diploma or the equivalent and who is absent from school for 30 or more consecutive days with no evidence of being enrolled in another public or private school" and authorized the Texas Education Agency to implement a system for collecting data on student dropouts (Keel, Alwin, and Nelson 2000). In 1987, HB 1010 significantly increased individual school districts' responsibility for collecting accurate dropout data and monitoring dropout rates. As the final component of the state's current mandates for educational organizations, Senate Bill (SB) 152 set statewide targets for dropout reduction and mandated that poor-performing districts formulate strategies and direct resources to reduce dropouts.[4]

With these three laws, the state government effectively incorporated dropout performance into the state's Academic Excellence Indicator System (AEIS). Designed as an accountability tool, the state uses AEIS scores to determine financial rewards to districts (TEC § 39.091, 1999), generate district performance reports (TEC § 39.053, 1999), and create school-level report cards for dissemination to parents (TEC § 39.052, 1999).

In addition to their own personal motivations for improving their organizations, Texas superintendents are required by state law to be interested in measuring dropout performance. Interestingly, however, the statute does not specify the *type* of dropout rate calculation that must be used. As a result of this statutory ambiguity, there are currently three agreed-upon measures: one used by the state education agency in its current assessment of districts, one recommended by the most recent gubernatorial commission on dropouts for inclusion in future assessments, and one used by the national government in its calculations of district-level dropout rates. We will briefly outline how each of these measures is calculated and discuss the challenges education managers face when they select among them.[5]

The most straightforward measure of dropouts—and organizational performance on the dropout "problem"—is the indicator used in the current AEIS accountability system. In 1987, the Texas Education Agency began calculating an annual dropout rate for

each district and school and reporting that figure to the legislature and the governor as required by law. The annual measure simply divides the number of students who drop out of grades 7–12 in a given year by the total number of students enrolled for that year. Proponents of the annual measure point to the fact that it is the most straightforward and least costly measure to calculate and that it can be disaggregated by grade level. Alternatively, critics argue that it produces the lowest figure of any calculation method and therefore may obscure poor organizational performance.

The U.S. Department of Education's National Center for Education Statistics (NCES) also calculates an annual dropout rate, but only for grades 9–12. The argument for dropping grades 7 and 8 from the calculation is that very few students choose to leave school during those grades, and thus their inclusion artificially deflates the actual dropout rate. Proponents of the NCES measure argue that it is a nationally agreed-upon indicator that can be used to compare performance across states. Critics suggest that the 9–12 annual measure still produces dramatically lower dropout numbers than other indicators. They also argue that it is highly unstable and prone to large annual fluctuations.

In response to many of the criticisms of both its annual indicator and the one used by the NCES, the Texas Education Agency began calculating a longitudinal dropout measure in 1997. The longitudinal indicator is a cohort measure that divides the number of students who drop out by grade 12 by the total number of students in the original ninth-grade class. This longitudinal measure has been suggested by scholars and measurement experts, and it is the one endorsed by a recent joint commission tasked with assessing the state's dropout problem. In December 2000, the Legislative Budget Board, the State Auditor's Office, and the Texas Education Agency presented a report to the 77th session of the Texas Legislature suggesting that the longitudinal dropout measure be incorporated as quickly as possible into the AEIS accountability standards (Keel, Alwin, and Nelson 2000). The report argued that the cohort measure is "more consistent with the public's understanding of a dropout" and more stable and reliable over time (20). It acknowledged, however, that the cohort measure requires considerably better data to calculate and cannot be disaggregated by grade.

So, the managers of Texas school districts are faced with three plausible measures of the same concept if they want to assess their organization's performance regarding dropouts. Each is used, or will soon be used, to hold districts accountable by some level of government. Interestingly, choosing the "purpose" for measurement does not instantly clarify the choice for

managers, as some authors have suggested (Behn 2003). For instance, a manager might want to measure dropouts to ensure that the district is performing at a level that will satisfy political principals. If the manager is most concerned with current state-level officials, he or she would clearly choose the 7th–12th-grade annual measure. In only a few short years, however, those same elected officials will be gauging district performance based on the 9th–12th-grade longitudinal measure. Perhaps, then, the farsighted manager should begin calculating that indicator. And finally, we must remember that the federal government is increasingly issuing calls for educational accountability in return for federal funds. It is basing its assessments on the 9th–12th-grade annual measure, and so the manager might also want to ensure that he or she is performing adequately by that standard.

Thus, the selection of performance measures for even a single purpose may be quite complex. Good managers, of course, want to measure performance not only to appease political principals but also to improve service for clients and for a host of other reasons—so the difficulty increases. As we have noted, the purpose of this study is not to offer advice about the selection of indicators but to illustrate how different acceptable measures may provide starkly divergent feedback to the manager regarding whether there is a problem, the causes of the problem, and the effectiveness of programmatic solutions. The following section will empirically explore the consequences of different performance measures using data from more than 711 Texas school districts.

Empirical Analyses of Differing Indicators of the Dropout Problem in Texas

The data for this study are drawn from 711 school districts in Texas. They are publicly available through the Texas Education Agency. In using these data to explore the potential impacts of multiple measures, our research strategy is quite straightforward. Simply put, we will test to see whether the different ways that managers might choose to measure dropouts lead them to different conclusions about whether their organizations are performing poorly, the primary causes of the dropout problem (if one exists), and the effectiveness of solutions designed to address that problem. In other words, we want to know whether a manager could assume that his or her district is performing considerably better than other districts using one indicator but find that it is actually well behind using another. Similarly, we want to know whether organizational features such as teacher experience and turnover are powerful predictors of dropouts with one measure but fail to predict the dropout rate when measured in different fashion.

In exploring these potential differences, we offer two analyses for the reader to consider. The first is the

most straightforward and simply ranks districts by their dropout rates, which are calculated using the three methods discussed previously, and plots the districts according to their rankings. The second analysis is an ordinary least squares regression in which we attempt to predict the dropout rate in Texas schools, calculated three different ways, with a set of widely agreed-upon organizational variables. Some of these variables are what the literature considers input characteristics, which are largely beyond the control of the manager. Others are institutional or policy-oriented measures, over which Texas superintendents have a considerable degree of influence.

Data The data for our analysis come from the Texas Education Agency and the National Center for Educational Statistics. Our sample includes all districts in Texas with at least 500 students for the 1999–2000 and 2000–01 school years.

Dependent Variables The dependent variables for our analysis are the three dropout measures discussed previously. To assess the relative differences across the measures, we created rankings for the average dropout rate across the two years of data. We use the actual dropout rate to assess the institutional and environmental impacts on these performance measures.

Independent Variables Our independent variables are designed to capture the environmental and institutional impacts on dropout rates. The environmental variables that influence dropout rates are factors that administrators have little or no control over. The first factor is district wealth. Although current finance-reform measures in the states have decreased financial inequities among districts, there is still a relationship between district wealth and district financial resources (Wood and Theobald 2003). To capture district wealth, we use the percentage of revenue generated from local sources. Logged district enrollment is used to capture the effect of district size. With increased district size come economies of scale, which could make greater resources available to deal with the dropout problem. However, larger districts tend to be located in urban areas, which could increase the dropout problem (Rumberger and Thomas 2000).

Finally, it has long been known that student characteristics are the best predictors of student performance (Coleman et al. 1966; Hanushek 1986). To capture these effects, we include the percentage of students who are classified as having limited English proficiency (LEP), the percentage of Latino and African American students, and the percentage of students who are classified as economically disadvantaged.

Institutional variables are factors that administrators have some influence over. The first institutional

variable is class size. The reduction of class size has long been advocated as a reform measure designed to improve school performance. Reduced class size affects both students and teachers by allowing students to receive more individual attention and improving teacher working conditions, both of which should lead to improved performance. Two recent studies have linked the reduction of class size to decreases in dropout rates (McNeal 1997; Rumberger and Thomas 2000).

Teacher quality has also been linked to dropout rates (Rumberger and Thomas 2000). Two institutional variables capture teacher quality: base teacher salary and teacher experience. Districts that offer higher salaries should be able to attract better teachers. Average teacher experience not only captures teacher experience but also a district's ability to keep teachers.

The final institutional variable used in this analysis is per-pupil instructional expenditures. There is a great deal of debate surrounding the efficacy of expenditures in improving school performance (e.g., Hanushek 1989; Hedges, Laine, and Greenwald 1994). Nevertheless, if schools use their financial resources efficiently, greater instructional expenditures should lead to decreased dropout rates.

Methods
Simple correlation analysis and scatter plots are used to assess the relative comparability between the three measures. To assess the institutional and environmental influences on the three dropout measures, we use pool time-series regression. We use a between-effects estimator that ignores within-district effects by estimating the average of the dependent variable for a district as a function of the average of each independent variable.[6] We use this estimator instead of a fixed effect estimator because we feel that the variance across the two years cannot be directly linked to variance across the years for the environmental and institutional variables. That is, the variance in dropout measures for a district from one year to the next is a function of random error instead of explanatory variables. This year-to-year variation is larger for small districts, which is what we would expect: Dropout rates in small districts are more sensitive to individual dropouts. The between-effects estimator essentially produces an average for the dependent variables over the two-year period, producing a measure that better represents true district performance, even for districts with low enrollment.

Findings and Discussion
Our research strategy assumes that a manager will first seek to understand whether his or her organization has a dropout problem relative to other school districts. Our assertion is that the different measures of this indicator that managers might plausibly choose give

them very different answers to this question. To determine whether this is the case, we must first get a general picture of the measures themselves. The cohort measure produces the highest mean dropout rate (5.39 percent), followed by the NCES dropout measure (3.4 percent); the AEIS measure produces the lowest mean (0.78 percent). To assess the relative comparability of these measures, we generated ranks for each district using the three measures. If they are essentially analogous, then the ranks across the measures should be highly correlated. The correlation between the cohort rank and the AEIS rank is the highest, with a correlation coefficient of 0.78, followed by the correlation between AEIS rank and NCES rank at 0.65, and 0.6 between the cohort and NCES rank.

Though these relatively low correlations between measures of the same concept are interesting, we are most interested in the information that each provides to managers about their organizations. If these measures provide similar information, then a district that is above (below) the median on one measure should, for the most part, be above (below) the median on the others. Figure 1 assesses whether this is the case by plotting districts based on the ranks assigned using the different scores. These scatter plots show the relationship between the ranks produced by the different measures. The horizontal and vertical lines that intersect the plots represent the median dropout rate for each measure. Observations in the upper-left and lower-right quadrants represent districts that are above the median on one measure and below the median on the other. For the relationship between the cohort and AEIS measures, 19 percent of the districts are above the median on one and below the median on the

other. In other words, one in five managers will get different answers about whether their organization has a problem depending on whether they choose the measure of dropouts currently used by the state education agency for purposes of evaluation or the measure that the agency will likely adopt in the near future.

For the other two relationships, the percentage of districts that lie above the median on one and below the median on the other is even higher. Again, these districts can be found in the upper-left and lower-right quadrants of the plots. Almost 30 percent of districts are above the 50th percentile on the cohort measure but in the bottom half of districts on the NCES indicator, and vice versa. In the comparison of the AEIS and NCES rankings, the percentage falls slightly to 27.5 percent. Together, these figures suggest that almost one in three managers may reach a different conclusion about the relative severity of the problem in their organization depending on the performance indicator chosen.[7]

The potential impact of different measures on managerial decision making in the aggregate is clearly visible in the comparisons just presented. It is easier to see the impact on individual managers, however, if we isolate a few school districts whose rank is strongly influenced by the choice of dropout measure. Lytle Independent School District, for example, is a relatively small district with a high percentage of Hispanic and economically disadvantaged students. The Lytle district recently received an "exemplary" rating from the Texas Education Agency based on its performance on the state high-stakes exam and the fact that, based on the AEIS indicator currently used by the state,
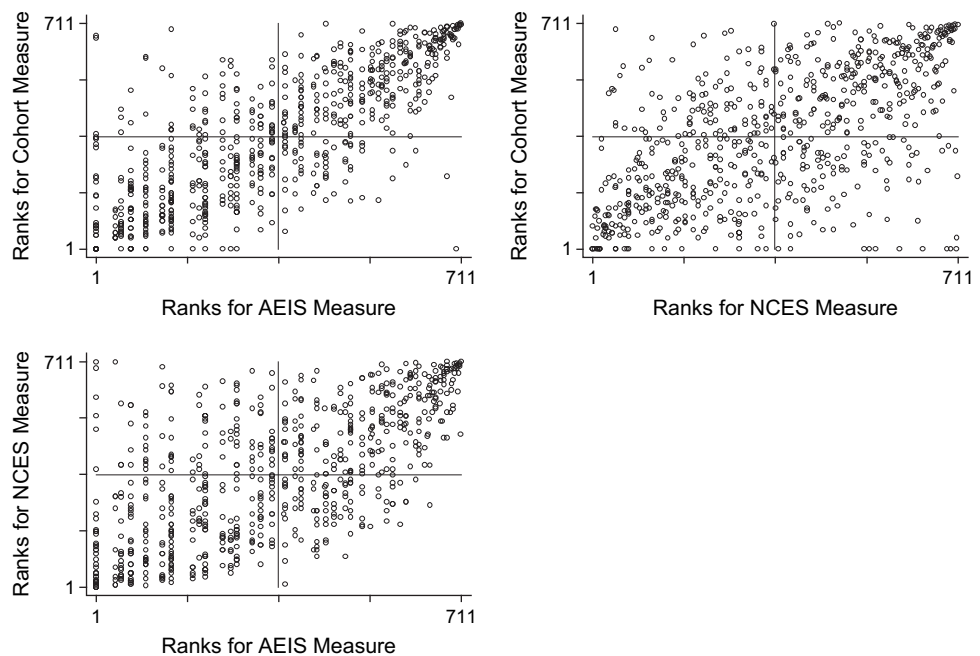

Figure 1   Relative District Rankings

only 65 districts have a lower dropout rate. Thus, this measure might indicate to the manager that little needs to be done about dropouts, particularly given the other challenges that the district faces resulting from its large number of economically disadvantaged students.

Interestingly, however, if the state switches to the proposed cohort measure in a year or two, the district's ranking will drop to 600th, which will certainly not qualify it for an exemplary rating. If, in anticipation of the upcoming change in state policy, the Lytle superintendent measured dropouts in his or her district using the cohort measure, he or she would likely conclude that significant resources needed to be diverted to dealing with the dropout problem.

Weimar Independent School District provides another interesting example of the disparity among dropout measures and the challenges facing superintendents who use them. Currently, this relatively poor and small district is ranked 577th in terms of dropout performance by the Texas Education Agency. If the state switches to the cohort measure, that ranking will rise to 263rd. Neither of these rankings is particularly impressive and might lead the superintendent to allocate resources to reducing dropouts rather than face potential sanctions from the state. When the federal Department of Education assesses Weimar's performance on dropouts, however, a very different picture emerges. Based on the measure calculated by the NCES, only 60 districts in the state have a lower dropout rate. Poorer districts are heavily dependent on federal funding, and thus, the superintendent of Weimar might have a real incentive to measure performance in the same fashion as federal regulators. Based on that measure, he or she might justifiably conclude there is little reason to devote scarce resources to dropout-reduction programs.[8]

We can now turn to the potential impact of different performance measures on managerial assessments of the causes of and solutions to organizational problems. Table 1 presents the findings from three models that predict differing dropout indicators with the environmental and institutional variables suggested by the literature on educational performance. The first column presents the regression results for the cohort measure, the second contains the findings from the AEIS indicator, and the third represents the model employing the NCES dropout measure. We have asserted thus far that the very different nature of these measures may lead to different findings about the causes and cures of the dropout problem.

The first variables that we wish to focus on are those that reflect the environment of the districts and the inputs that managers of these organizations have to work with. First, these models offer no evidence that district wealth, as measured by the percentage of revenue generated by local resources, affects dropout rates. The next environmental indicator that we consider reflects the number of clients—in this case, students—served by the organization. Here managers might draw very different conclusions from analyses of different dropout measures. Our indicator of enrollment is a significant (one-tailed) predictor of dropouts using both the cohort and the NCES measures. In both cases, the findings suggest that larger schools can expect higher dropout rates, with a 10 percent increase in enrollment translating to a roughly 1 standard deviation increase in the rate. Alternatively, the measure fails to significantly predict dropouts in the model of the AEIS annual measure. Thus, one measure tells a superintendent expecting rapid growth within the district, perhaps because of the spread of suburbs, to anticipate a growing dropout problem. Alternatively, the other measures tell that same superintendent that he or she should not focus on dropouts, but on the other problems that often accompany rapid growth.

The next set of environmental variables capture student characteristics. The literature on dropouts suggests that students with language difficulties are more likely to drop out and that officials in districts with a high percentage of these students must be particularly attentive to the dropout problem. The percentage of students classified as LEP, however, is only a significant predictor of dropouts in the NCES model. In addition, it is in the unexpected direction. That is, as this measure increases, dropouts decrease. This could be a function, though, of LEP students dropping out before the grades that are measured for each of the dropout rates. This might also explain why this measure is not significantly related to the other two dropout measures. The inconsistent results for this measure highlight not only the differences between the measures but also the limitations of all of these measures to actually capture the true nature of the dropout problem.

Indicators of the percentage of at-risk students, particularly those that reflect the economic status of clients, also appear to have disparate impacts depending on the performance measure selected. In this case, the measure is significant in all three models, but the substantive impact is dramatically different in each. Adjusting for the differences in the standard deviations of these measures, the percentage of low-income students within a district has a 53 percent larger impact on the cohort indicator compared to the AEIS measure. Thus, a manager who sees the results from a model predicting the Texas Education Agency's current measure of dropouts would be significantly less concerned about a predicted increase in low-income students than one who sees evaluations using the cohort measure.

Table 1 Regressions for Three Dropout Measures

	Cohort Measure	AEIS Measure	NCES Measure
Environmental Variables			
Percentage local revenue	−.002	.000	.004
	(.27)	(.14)	(.78)
Logged enrollment	.495	.046	.193
	(3.03)**	(1.37)	(1.74)
Percentage LEP	.015	−.004	−.030
	(.74)	(1.06)	(2.26)*
Percentage low-income	.056	.006	.041
	(4.05)**	(1.99)*	(4.41)**
Percentage African American	.024	.007	.018
	(1.82)	(2.78)**	(2.06)*
Percent Latino	.017	.007	.016
	(1.68)	(3.46)**	(2.42)*
Institutional Variables			
Class size	−.194	.013	.277
	(2.15)*	(.68)	(4.51)**
Teacher salary	−.000	−.000	−.000
	(.10)	(.13)	(.79)
Teacher experience	.031	−.010	−.096
	(.46)	(.73)	(2.12)*
Instructional expenditures	−.001	−.000	−.000
	(3.38)**	(1.53)	(.45)
Constant	4.902	.306	−1.656
	(1.83)	(.56)	(.91)
Observations	1,392	1,392	1,392
Number of districts	711	711	711
R^2	.24	.17	.27

Note: Absolute value of *t* statistics appears in parentheses.
*significant at 5 percent; **significant at 1 percent.

The environmental variables are interesting because they can inform superintendents about the causes of the dropout problem in their organizations. Additionally, they are important for the purposes of this study because the results suggest that managers' perceptions of their importance may vary dramatically based on the choice of measure. Nonetheless, environmental variables cannot be of much help to managers in crafting solutions to the dropout problem because normal educational managers do not have a great deal of control over them. There are a number of organizational features and policies, however, over which superintendents do exercise significant influence.

The educational performance literature suggests that one of the best things superintendents can do to reduce dropouts is to shrink student–teacher ratios. As the findings in table 1 suggest, however, education managers may reach very different conclusions about the effect of class size depending on the measure of dropouts they choose. The annual measure currently favored by the Texas Education Agency suggests there is no statistical relationship between dropouts and class size. The NCES measure, on the other hand, suggests that superintendents may be able to reduce dropouts by a full standard deviation simply by reduc-

ing the number of students in the average class by one. This indicates that expenditures on new teacher hires would be money well spent. Interestingly, however, managers who chose to measure dropouts using the cohort measure might reach exactly the opposite conclusion. The negative coefficient on the class size variable in the first column suggests that as class size increases, the dropout rate will decrease. This is counterintuitive and goes against the predictions in much of the literature, and so managers would probably be hesitant to start firing teachers to reduce dropouts. Nonetheless, the finding might be expected to significantly complicate decision making about effective solutions to the dropout problem.

The next organizational and policy variables that we want to discuss relate to teacher quality. The education literature suggests that personnel quality may influence dropout rates and acknowledges that recruitment of such personnel is an area in which superintendents play a large role. In terms of teacher experience (a key indicator of quality), we again find that different measures of dropouts provide different feedback. For both the cohort measure and the AEIS annual indicator, the insignificant coefficient for teacher experience suggests there is no relationship

between this feature of personnel and dropouts. Alternatively, the NCES annual measure indicates that more experienced teachers are more able to keep kids in school. Hiring and maintaining an experienced workforce is expensive, and once again, the way that superintendents choose to measure dropouts is likely to have an impact on their assessments of whether this is money well spent.

Finally, the results regarding instructional expenditures offer another interesting example of the potential impact of different measures. One of the longest-raging debates in education centers on whether spending more on classroom instruction leads to better performance. This is particularly true in the case of dropouts, for whom nonclassroom activities such as sports, clubs, and after-school activities are often advanced as preferred solutions. According to our findings, a superintendent's conclusions regarding this debate might well depend on how he or she chooses to measure dropouts. For both the state (AEIS) and national (NCES) annual measures, there is no significant relationship between instructional expenditures and dropout rates. This might lead superintendents to conclude that money allocated to reducing dropouts might best be spent elsewhere. Alternatively, the model of the cohort measure does suggest a significant relationship and could suggest to education managers that spending more in the classroom is a workable solution to the dropout problem in their districts.

Conclusion
We assert that public managers often face many suitable performance measures of the same concept and that their choice of measure can have a real impact on conclusions about organizational performance. To illustrate this point, we have drawn on a case from public education in the state of Texas, where superintendents are faced with three measures of student dropouts, each of which is endorsed by a set of political principals. The findings from our empirical analysis of school districts in Texas suggest that each of these measures may produce a starkly different conclusion concerning (1) whether an organization has a dropout problem, (2) the causes of that problem, and (3) the effectiveness of organizational solutions to the problem. We believe these findings offer considerable support for our assertion that differing measures can provide starkly different feedback to managers about their organizations.

The results also lead us to two related conclusions. The first is that scholars need to treat the complexities introduced by multiple measures more carefully if their research is to meaningfully inform the practice of performance measurement in public organizations. Research on the purposes and types of performance measurement being used in public organizations, as well as the benefits of such activities, is undeniably useful. However, this research may also oversimplify the choices faced by managers who want to measure performance. Future research should acknowledge the difficulty of choosing among multiple measures and develop criteria that can assist managers in those choices. One criterion might be the match between measurement strategy and organizational characteristics. The different measures explored in this study were particularly sensitive to specific changes in the organization: The cohort measure was most unreliable in very small districts, and the AEIS measure produced misleading results for districts that had experienced high enrollment growth or decline in the lower grades. We imagine that there are many instances in which the match between organizational characteristics and measurement strategy can help determine the usefulness of the information provided to managers.

The second conclusion that we draw from this research concerns the development, maintenance, and adaptation of performance-management systems. In a federal system, in which multiple levels of government may hold public organizations accountable for their performance, it is highly counterproductive for each set of political principals to measure performance differently. It is counterproductive because when districts such as Splendora Independent School District rank in the top 10 percent on the state-level measure of performance and in the bottom 50 percent on the federal measure, there is no consistent guidance for managers concerning the allocation of scarce resources. It is also counterproductive because these types of disparities are likely to encourage gamesmanship such as that observed in job-training programs (Blau 1955), in which managers search for the most flattering measurement strategy rather than improve performance.

In addition to the obvious problems of maintaining performance-measurement systems with conflicting measures, the analyses presented here also suggest that governments need to be careful when substituting new measures for old. Nowhere in the report to the Texas Legislature on the superiority of the cohort measure did the Texas Education Agency discuss the ramifications of radically changing the measurement strategy for dropout performance. Remember districts such as Lytle Independent School District, which drop from the top 5 percent on dropout performance to the bottom 50 percent when the change is implemented, even if not a single additional kid drops out. Organizational performance is increasingly being tied to public funding and other benefits, and so the designers of public accountability standards must be sensitive to the real differences among multiple measures of performance. Additionally, governments must develop strategies for dealing with organizations whose perceived performance radically changes with the adoption of new measures.

Notes

1. Kravchuk and Schack (1996) suggest that the real challenge comes in aggregating the individual measures into some usable index of overall performance.

2. The No Child Left Behind program is the most recent example of federal accountability proposals; some form of measurement and reporting system exists in all 50 states.

3. Financing is typically dominated by local district tax revenues, along with state funding. The national government provides relatively little public education financing in most districts.

4. SB 152 originally passed in 1989 and was reauthorized in 1995.

5. Any study claiming that others have understated the complexities of a particular issue needs to be honest about the way it also fails to address all aspects of that issue. For example, in the case of dropouts, we suggest that managers use one of three summary measures to (1) determine whether a problem exists, (2) devise a solution, and (3) assess the effectiveness of that solution. This is obviously an oversimplification of the process of program assessment that managers must undertake. A recent pilot study of South Texas school districts, conducted jointly by Texas A&M University and the University of Texas–Pan American, found that some districts were operating as many as 18 dropout-prevention programs simultaneously (survey results are available from authors). Clearly, the way that managers use available dropout measures to assess the relative or total effectiveness of these programs is a more complex topic than we can adequately treat in the space available here. In the interest of parsimony, this manuscript also ignores other elements that contribute to the challenges of performance measurement. Namely, we focus on one output to the exclusion of all others. In other words, we look at dropouts but not at standardized test performance, attendance, athletics, bilingual and vocational education, or any of the other metrics on which school districts are simultaneously judged. Operating as boundedly rational decision makers in a complex organization, superintendents' decisions regarding which dropout measure to choose and what weight to put on subsequent findings are undoubtedly influenced by the performance and expectations on these other dimensions.

6. Instead of estimating $y_{it} = \alpha + \mathbf{x}_{it}\beta + \upsilon_i + \varepsilon_{it}$, a between-effects estimator estimates $\bar{y}_i = \alpha + \bar{\mathbf{x}}_i\beta + \upsilon_i + \bar{\varepsilon}_i$.

7. The disparate rankings produced by each measure give little reason for concern if districts are all clustered just above and just below the median; managers are surely savvy enough to know that there is no substantive difference between the 48th and the 52nd percentile. However, as the scatter plots in figure 1 suggest, this is not the case. In fact, between the cohort and AEIS measures, 30 districts (4.2 percent) are ranked in the top third on one and the bottom third on the other. Between the cohort and NCES measures, 34 districts (4.8 percent) are ranked in the top third on one and the bottom third on the other. Finally, between the AEIS and NCES measures, 36 districts (5.1 percent) are ranked in the top third on one and the bottom third on the other. Thus, the findings regarding disparate feedback cannot simply be dismissed as an artifact of the arbitrary divide between the top and bottom halves in a summary statistic.

8. The differences in rankings produced by disparate measures of performance may be attributable to factors other than the actual number of kids who drop out of school. For instance, differences between the AEIS and the NCES measures can be produced by different distributions of students across grade levels. Thus, to fully understand the feedback produced by any measure, managers must be able to (1) break out its component elements and (2) compare those with the factors used to calculate other measures. In keeping with our general argument in this article, this adds still more complexity for managers who want to measure performance meaningfully in their organizations.

References

Altshuler, Alan. 1988. A Comment on Groping Along. *Journal of Policy Analysis and Management* 7(4): 664–67.

Ammons, David. 1992. Productivity Barriers in the Public Sector. In *The Public Productivity Handbook*, edited by Mark Holzer, 117–36. New York: Marcel Dekker.

———. 1995. Overcoming the Inadequacies of Performance Measurement in Government: The Case of Libraries and Leisure Services. *Public Administration Review* 55(1): 37–52.

Barnett, W. Steven. 1996. Economics of School Reform: Three Promising Approaches. In *Holding Schools Accountable: Performance Based Reform in Education*, edited by Helen Ladd, 299–326. Washington, DC: Brookings Institution Press.

Behn, Robert. 2003. Why Measure Performance? Different Purposes Require Different Measures. *Public Administration Review* 63(5): 586–604.

Berman, Evan, and XiaoHu Wang. 2000. Performance Measurement in U.S. Counties: Capacity for Reform. *Public Administration Review* 60(5): 409–20.

Bishop, John. 1994. The Impact of Curriculum-Based Examinations on Learning in Canadian Secondary Schools. Working Paper No. 94, Cornell University, Department of Economics.

Blau, Peter. 1955. *The Dynamics of Bureaucracy: A Study of Interpersonal Relations in Two Government Agencies.* Chicago: University of Chicago Press.

Bouckaert, Geert. 1992. Public Productivity in Retrospective. In *The Public Productivity Handbook*, edited by Mark Holzer, 5–46. New York: Marcel Dekker.

Brandon, P., and R. Heck. 1998. The Use of Teacher Expertise in Decision-Making During School Conducted Needs Assessments: A Multilevel Perspective. *Evaluation and Program Planning* 21(3): 323–31.

Cohen, Steven, and William Eimicke. 1995. *The New Effective Public Manager*. San Francisco: Jossey-Bass.

Coleman, James S., et al. 1966. *Equality of Educational Opportunity*. Washington, DC: U.S. Department of Health, Education, and Welfare.

de Lancer Julnes, Patria, and Marc Holzer. 2001. Promoting the Utilization of Performance Measures in Public Organizations: An Empirical Study of Factors Affecting Adoption and Implementation. *Public Administration Review* 61(6): 693–708.

Drucker Peter. 1973. *Management*. New York: Harper & Row.

Duncomb, William, John Ruggiero, and John Yinger. 1996. Alternative Approaches to Measuring the Cost of Education. In *Holding Schools Accountable: Performance Based Reform in Education*, edited by Helen Ladd, 327–56. Washington, DC: Brookings Institution Press.

Ekstrom, Ruth B. 1986. Who Drops Out of High School and Why? Findings from a National Study. *Teachers College Record* 13(2): 356–73.

Fletcher, Michael A. 2002. Exit Tests Hurt At-Risk Students. *Washington Post*, August 14.

Hanushek, Eric A. 1986. The Economics of Schooling: Production and Efficiency in Public Schools. *Journal of Economic Literature* 24(3): 1141–71.

———. 1989. The Impact of Differential Expenditures on School Performance. *Educational Researcher* 18(1): 45–61.

Haveman, Robert, and Barbara Wolfe. 1995. The Determinants of Children's Attainments: A Review of Methods and Findings. *Journal of Economic Literature* 33(4): 1829–78.

Hatry, Harry. 1999. *Performance Measurement: Getting Results*. Washington, DC: Urban Institute.

Hedges, Larry V., Richard D. Laine, and Rob Greenwald. 1994. Does Money Matter? A Meta-Analysis of Studies of the Effects of Differential School Inputs on Student Outcomes. *Educational Researcher* 23(1): 5–14.

Heinrich, Carolyn. 2002. Outcomes-Based Performance Management in the Public Sector: Implications for Government Accountability and Effectiveness. *Public Administration Review* 62(5): 712–25.

Keel, John, Lawrence Alwin, and Jim Nelson. 2000. Dropout Study: A Report to the 77th Texas Legislature. Houston: Legislative Budget Board, State Auditors Office, and Texas Education Agency.

Kopczynski, Mary, and Michael Lombardo. 1999. Comparative Performance Measurement: Insights and Lessons Learned from a Consortium Effort. *Public Administration Review* 59(2): 124–34.

Kravchuk, Robert, and Ronald Schack. 1996. Designing Effective Performance Measurement Systems under the Government Performance and Results Act of 1993. *Public Administration Review* 39(2): 348–58.

Ladd, Helen, ed. 1996. *Holding Schools Accountable: Performance Based Reform in Education*. Washington, DC: Brookings Institution Press.

McNeal, Ralph B., Jr. 1997. High School Dropouts: A Closer Examination of School Effects. *Social Science Quarterly* 78(2): 209–22.

Meyer, Robert. 1996. Comments on Chapters Two, Three, and Four. In *Holding Schools Accountable: Performance Based Reform in Education*, edited by Helen Ladd, 137–45. Washington, DC: Brookings Institution Press.

Milgrom, Paul, and John Roberts. 1992. *Economics, Organizations, and Management*. New York: Prentice Hall.

Murphy, David. 1999. Presenting Community-Level Data in an "Outcomes and Indicators" Framework. *Public Administration Review* 59(1): 76–82.

Newcomer, Kathryn. 1997. Using Performance Measures to Improve Programs. *New Directions for Program Evaluation* 75(1): 5–14.

Nicholson-Crotty, Sean, and Kenneth J. Meier. 2003. Size Doesn't Matter: In Defense of the Single State Study. *State Politics and Policy Quarterly* 2(4): 411–22.

Osborne, David, and Ted Gaebler. 1992. *Reinventing Government. How the Entrepreneurial Spirit Is Transforming the Public Sector*. Reading, MA: Addison-Wesley.

Osborne, David, and Peter Plastrik. 2000. *The Reinventor's Fieldbook: Tools for Transforming Your Government*. San Francisco: Jossey-Bass.

Orr, M. T. 1987. *What to Do about Youth Dropouts? A Summary of Solutions*. New York: Structured Employment/Economic Development Corporation.

Phillips, Meredith, James Crouse, and John Ralph. 1998. Does the Black-White Test Score Gap Widen after Children Enter School? In *The Black-White Test Score Gap*, edited Christopher Jencks and Meredith Phillips, 229–72. Washington, DC: Brookings Institution Press.

Rumberger, Russell W., and Scott L. Thomas. 2000. The Distribution of Dropout and Turnover Rates among Urban and Suburban High Schools. *Sociology of Education* 73(1): 39–67.

Thompson, James. 1967. *Organizations in Action*. New York: McGraw-Hill.

Velez, William. 1989. High School Attrition among Hispanic and Non-Hispanic White Youths. *Sociology of Education* 62(2): 119–33.

Wholley, Joe. 1999. Performance-Based Management: Responding to the Challenges. *Public Productivity and Management Review* 22(3): 288–307.

Wood, B. Dan, and Nick A. Theobald. 2003. Political Responsiveness and Equity in Public Education Finance. *Journal of Politics* 56(3): 718–38.

Worthington, A., and B. Dollery. 2000. Measuring Efficiency in Local Governments' Planning and Regulatory Function. *Public Productivity and Management Review* 23(4): 469–85.

We Invite Your Feedback

The **PAR** editorial team is in the process of designing an interactive, web-based accessory to the journal to facilitate dialogue and exchange on the printed content.

In the meantime, we invite thoughtful comments in the form of letters. E-mail to: par.letters@cudenver.edu.

Conventional mail: Public Administration Review, GSPA, University of Colorado at Denver and Health Sciences Center, 1380 Lawrence Street, Suite 500, Denver, CO 80204.

Selected letters will be reproduced in future issues!