Dual standards of school performance and funding? Empirical searches of school funding adequacy in Kentucky and Maine

Jaekyung Lee

* Graduate School of Education, State University of New York at Buffalo, Buffalo, New York, USA

First Published on: 25 March 2009

To cite this Article Lee, Jaekyung(2009)'Dual standards of school performance and funding? Empirical searches of school funding adequacy in Kentucky and Maine', Education Economics, 99999:1, To link to this Article: DOI: 10.1080/09645290902796415, URL: http://dx.doi.org/10.1080/09645290902796415
Dual standards of school performance and funding? Empirical searches of school funding adequacy in Kentucky and Maine

Jaekyung Lee*

Graduate School of Education, State University of New York at Buffalo, Buffalo, New York, USA

This study examines potential consequences of the discrepancies between national and state performance standards for school funding in Kentucky and Maine. Applying the successful schools observation method and cost function analysis method to integrated data-sets that match schools’ eight-grade mathematics test performance measures to district funding, this study conducts empirical searches for adequate instructional expenditures per pupil to meet desired proficiency targets on national versus state assessments. While Kentucky (high-stakes testing state) had a lower performance standard than Maine (low-stakes testing state), this study reveals a relatively lower level of school funding adequacy and a weaker relationship between school expenditures and performance for Kentucky than for Maine. The study suggests that state educational accountability systems and policies may influence the level of state performance standards and the proficiency gaps between national and state assessments, which in turn lead to potential gaps in school funding. Implications for policy and research are discussed to address problems with dual standards of school performance and to improve school funding adequacy and efficiency.

Keywords: accountability; adequacy; school funding; performance standards; high-stakes testing

Introduction

The No Child Left Behind Act of 2001 (NCLB), aimed at high academic standards for all students, requires that schools have all students become ‘proficient’ in reading and mathematics by the 2013/14 school year. The definition of proficiency is determined by individual states; it is based on their own performance standards, and is usually measured by the states’ own assessments. Although the NCLB does not prescribe the role of National Assessment of Educational Progress (NAEP), the nation’s report card, for state educational accountability systems, the expected use of NAEP was to confirm state test results under the purview of the US Department of Education (Henderson-Montero, Julian, and Yen 2003; Kentucky Department of Education 2004).

Despite potential utility of NAEP as a validation tool, previous comparisons of NAEP and state assessment results showed significant discrepancies in the level of student achievement, as well as in the size of statewide achievement gains (Fuller et al. 2006; Klein et al. 2000; Koretz and Barron 1998; Lee 2006, 2007; Linn, Baker, and Betebenner 2002). While previous studies examined the discrepancies between

*Email: JL224@buffalo.edu
NAEP and state assessment results at the state level, none examined the same problem at the school level. As the target of current federal and state accountability policies focuses on individual schools, it becomes more crucial to use multiple measures for school accountability and examine discrepancies between different assessment results at this level. There can be variations among schools within the same state because schools may align their curriculum and instruction with different assessments (national, state, or district) to varying degrees.

Further, the issue of a gap between national and state performance standards has implications for school funding. As expectations for schools rise and money remains tight, policy-makers face growing pressures to strengthen the connection between school funding policies and academic results (Quality Counts 2005). In the 1990s, the focus of school finance reform shifted from equity to adequacy in the midst of the performance-based educational accountability movement (Bartman 2002; Clune 1994; Ladd and Hansen 1999). However, states that adopted high-stakes testing policies during the past decade often failed to improve key school resources and address funding inequalities (Lee and Wong 2004; Powers 2004). One of the most critical policy questions is: Do schools have sufficient resources to meet imposed proficiency targets? This becomes a more pressing issue as many disadvantaged, lower-achieving school districts and schools nationwide fail to meet Adequate Yearly Progress (AYP) performance targets and face potentially serious consequences under NCLB.

Recently, there have been attempts to estimate the cost of raising all children’s test scores to a particular state’s standard. Although different studies using different methods with data from different states suggested different cost estimates, most revealed needs for massive new investments in education spending; increases in base cost were in the 15–46% range (Mathis 2003; Rebell 2006). All of the previous studies relied solely on state assessment results when they determined the level of adequate school funding relative to state performance standards. However, the results could have been quite different if they had used the NAEP assessment, which usually employs much higher performance standards than their own state assessment. As Congress moves to reauthorize the NCLB, it is poised to discuss the topic of increasing the rigor of state standards and tests by linking them to those set at the national level (Olson 2007). However, it is not clear whether states can be induced to adopt more rigorous performance standards under the NCLB, unless the federal government substantially increases the level of its financial support to help them implement higher standards.

In light of these concerns, the purpose of this study is to investigate the discrepancy between NAEP and state assessment results on school performance and to draw inferences about the adequacy of school funding based on performance standards for the NAEP versus state assessments. Specifically, this study examines three research questions. First, how well do schools perform relative to NAEP and state performance standards? How large are the discrepancies between NAEP and state assessments in their estimation of the percentage of proficient students? Second, what are the adequate levels of funding that would enable schools to reach desirable levels of NAEP and state performance standards? How large are the discrepancies between NAEP-based and state assessment-based estimations of school funding adequacy? Third, what are the implications of dual standards of school performance for funding the implementation of NCLB? This study does not investigate the causes of the discrepancies between NAEP and state assessment
results, but instead focuses on examining the consequences of these discrepancies for funding issues.

**Conceptual and analytical framework**

This study is not intended to judge whether NAEP or a state assessment is more or less appropriate for evaluating academic performance of students and their schools. Rather, the study builds upon the assumption that NAEP is the only national assessment based on common curricular framework and performance standards, and thus it can be used for cross-checking different states’ achievement test results in the same subject for the same grade. Differences between NAEP and state assessments in their purpose and performance standards have been noted and their comparability has been debated (see Koretz 2003; Linn 2000; Loveless 2006). As the NCLB requires using ‘proficient’ as its target achievement level, the appropriateness of current NAEP achievement levels, particularly proficiency level as a benchmark for states, also remains questionable (see Shepard et al. 1993; US General Accounting Office 1993). The NAEP achievement levels are authorized by the NAEP legislation and adopted by the National Assessment Governing Board. They are collective judgments, gathered from a broadly representative panel of teachers, education specialists, and members of the general public, about what students should know and be able to do relative to a body of content reflected in the NAEP assessment frameworks.

This study is also based on the assumption that school expenditures and resources matter in improving academic achievement across the nation and states. The study attempts to address the question of what level of instructional resources is sufficient to reach national versus state performance standards. Syntheses of a large body of research, ‘education production function’ studies, provided mixed evidence on the effects of educational expenditures on student achievement (Hanushek 1997; Hedges, Laine, and Greenwald 1994; Monk 1992). Hanushek (1994) argued that the lack of a systematic relationship between school resources and student achievement makes a definition of ‘adequacy’ in funding and resources virtually impossible. However, some studies documented the effects of school resources such as instructional spending and well-qualified teachers on academic achievement (Ferguson 1991; Ferguson and Ladd 1996). Previous studies linking NAEP student achievement to school resources also showed indirect, positive relationships (Lee 2004; Wenglinsky 1998). District-level instructional expenditures are expected to affect school-level resources, which in turn affect student-level academic achievement. Research has shown that expenditures are typically invested in one of two resources, increasing the number of teachers per student or improving teacher quality (Odden and Clune 1995). Therefore, higher per-pupil expenditures (PPE) are likely to lead to smaller classes (with more teachers per student) or better-qualified teachers (with higher teacher salaries and more professional development).

Even when it is assumed that there is a significant relationship between key school resources and performance, there is still debate on how to determine an adequate level of school funding relative to state performance standards (Quality Counts 2005). While there are several approaches to measuring school funding adequacy, every approach has limitations and thus the results should be interpreted cautiously. An empirical observation of successful districts/schools strategy has been applied to Ohio and Illinois (Guthrie and Rothstein 1999). This strategy – the so-called ‘successful schools’ approach hereafter – determines a level of student performance considered
adequate, and then identifies school districts or schools that achieve the desired goals. The level of resources expended by such school districts or schools is then deemed to be adequate. However, this empirical approach has some drawbacks (see Guthrie and Rothstein 1999; Reschovsky and Imazeki 2001; Verstegen 2002). In particular, it does not explicitly take into account differences among schools in their student characteristics. The same proficiency measure is applied to schools based on the presumption that, regardless of racial and socio-economic status, all schools must achieve the same average proficiency in order to be considered to have adequate outcomes.

A more sophisticated econometric approach to estimating adequacy with better adjustments was applied to New York data by Duncombe, Ruggiero, and Yinger (1996), to Texas data by Reschovsky and Imazeki (2003, 2004), and to Wisconsin data by Reschovsky and Imazeki (1998). Their strategy employs cost function analysis to predict hypothetical spending for each district in order to achieve desired level of outcome. Then, this foundation level of adequacy can be adjusted for regional cost of living differences and for student poverty rates. Critics of this method claims that it is too complex to explain to policy-makers and also that the precision implied by statistical modeling may be misleading because underlying assumptions and judgments made in the data analyses are not necessarily without flaws (Guthrie and Rothstein 1999). For example, in Texas – where over 300 school districts challenged the constitutionality of the state system of school finance in 2004 – two cost function analyses conducted by different researchers (Imazeki and Reschovsky [2004] for the plaintiff districts versus Taylor [2004] for the State of Texas) reached opposite conclusions about the sufficiency of school district resources for meeting the state’s accountability standards; the discrepancy was attributed to their different underlying assumptions (Imazeki and Reschovsky 2005).

Using different measures of school performance or applying different methods to derive school funding adequacy may lead to different results and conclusions. In fact, it is not that one particular measure or method is always superior to the others, but that each measure or method has its own utilities and limitations. On one hand, using multiple measures and applying multiple methods may help enhance the validity of research to estimate school funding adequacy. On the other hand, using uncommon measures of performance runs the risk of producing incompatible results and ultimately dual standards of performance and funding. This study addresses these issues through empirical tests. Figure 1 illustrates an analytical framework for this study, tracking the influences of discrepancies between NAEP and state assessments (performance standards) on the gap in school performance proficiency, which in turn leads to the gap in school funding adequacy.

Data and methods
In this study, two states – Kentucky and Maine – have been selected purposively. The comparison of Kentucky and Maine provides us with contrast of school performance, accountability and funding policies. Early on, both states (a) established student assessment systems to monitor their schools’ academic progress and (b) made a greater effort to align their assessments with their content and performance standards. Despite these common characteristics, the two states’ larger educational contexts and accountability policies differed significantly (Lee 2007). According to the NAEP results, Kentucky is a relatively lower-achieving state while Maine performs at the top
in the nation. The Kentucky Instructional Results Information System (KIRIS) was designed to measure school improvement and to give schools rewards or sanctions based on the test results. Kentucky’s earlier model of academic improvement (toward the goal of 100% of students proficient over 20 years) was similar to NCLB’s AYP. Not as high-stakes a test as KIRIS, Maine Educational Assessment (MEA) was designed primarily to provide information to schools to assist in making decisions about curricula and instruction. According to the classifications of states by prior research, Kentucky is a strong accountability states with high-stakes testing, whereas Maine is a weak accountability state with low-stakes testing (Carnoy and Loeb 2002; Lee and Wong 2004).

The percentages of students reaching the ‘proficient’ level on NAEP tend to be generally lower than on the state assessments. Figure 2 shows 1996 NAEP and state assessment eighth-grade mathematics proficiency rates for 12 states with matched test results available. These results have been interpreted as implying that, for many states, NAEP proficiency levels are more challenging than the states’ own, and that state standards are still not high enough (see National Education Goals Panel 1996). Among the states included in Figure 2, Maine was one of the two states that showed an exception to this general pattern. Other studies found that similar patterns of discrepancies between national and state assessment results were shown in reading as well, and the same problems tend to continue after the NCLB (see Fuller et al. 2006; Lee 2006, 2007).

The policy context of the two states’ school funding also facilitates comparison. The trend toward adequacy litigation began with the 1989 landmark decision by the
Kentucky Supreme Court, which declared the state’s entire school system inadequate and led to school finance reform during the following decade; Kentucky has achieved greater school funding equity and increased support for its common school through reform with a new school funding formula including a base guarantee per pupil and required local tax effort (Adams and White 1997; Lefkowits 2004). In contrast, Maine’s school funding formula remained relatively constant during the past decade until the state adopted adequacy-driven school funding in 1999 to support its Learning Results, the state’s content and performance standards (National Center for Education Statistics [NCES] 2001). While the legislature directed the State Board of Education to develop a method of funding the programs and services necessary for students to achieve the statewide learning standards, it took the first steps in increasing the state subsidy and the per-pupil guarantee (Silvernail 1998, 2000). In comparison with Kentucky, Maine had relatively higher property value and lower poverty rate, which resulted in higher spending on public education. Education spending per pupil in 1998, adjusted for regional cost differences, was $6196 in Kentucky and $6739 in Maine (Quality Counts 2000).

Specifically, I utilized data collected from the states’ student assessments; that is, KIRIS and MEA in mathematics at Grade Eight in 1996. I also used the 1996 NAEP
eighth-grade mathematics state assessment data for Maine and Kentucky. The original data consisted of 312 public schools in Kentucky and 227 public schools in Maine. Among those schools, only a subset of schools that participated in both national and state assessments were used for analysis; the number of eligible schools was 90 in Kentucky and 92 in Maine (see Appendix 1 for the profiles of analytical samples). Unlike the state assessments that were given to all eligible students, the NAEP test was taken only by a random sample of students drawn from a set of randomly selected schools in each state. Because NAEP did not provide identifying information on individual schools that participated in the assessment, I used restricted NAEP data in tandem with Common Core of Data for matching NAEP schools to schools in the state assessment data-sets through common school identification codes. Because NAEP employed a multistage stratified random sampling method, its sample was designed to properly represent all public schools and their students in each participating state (see Appendix 1, Table A1). Therefore, the analysis utilized information from an overlapping subset of public schools that participated in both NAEP and state assessments and were presumably representative of the states’ target population. As Table A1 in Appendix 1 shows, although study samples in the two states are highly homogeneous and comparable with each other, they consist of predominantly small rural schools with relatively low percentages of minority students so that the study results may not be generalized to other states with different settings.

For each school in the analytic sample, the percentage of students at or above the Proficient level was calculated separately with NAEP, KIRIS, and MEA. The comparability of NAEP and the state assessment could be less problematic in Maine and Kentucky than other states because they modeled their frameworks closely after NAEP and adopted very challenging performance standards. Both Maine and Kentucky have achievement levels that are very similar to the NAEP levels. In Maine, proficiency levels were introduced into the MEA in 1995, and students were identified as being at Novice, Basic, Advanced, or Distinguished levels of achievement. In Kentucky, four corresponding categories were established for the KIRIS in 1992: Novice, Apprentice, Proficient, and Distinguished. While Kentucky set its student performance goal at the Proficient level on the KIRIS as a result of statewide education reform, Maine did not specifically link their performance standards with the MEA proficiency levels that lacked the ‘Proficient’ label. Despite the lack of a clear standards-assessment linkage as of 1996, it was reasonable to assume that Maine set its performance expectation for all students to the level of being ‘Advanced’ (second to the highest achievement level) on the MEA.2

Although the current NAEP cannot be used to measure and report academic proficiency of individual schools for accountability purposes, it has potential to produce sufficiently reliable school-level estimates of proficiency for research purposes. NAEP used item response theory to estimate proficiency scores in mathematics for each individual student. Rather than having a single observed mathematics score, there is a range or distribution of plausible values for each sampled student’s proficiency in NAEP mathematics. Plausible values were developed as a computational approximation to obtain consistent estimates of population characteristics in assessment situations where individuals are administered too few items to allow precise estimates of their ability (see Mislevy 1991; Mislevy, Johnson, and Muraki 1992). There are five such plausible values resulting from five random draws from the conditional distribution of proficiency scores for each student. Thus, the estimation of NAEP proficiency rate – that is, the percentage of students at or above Proficient
level – was based on the average of proficiency estimates from separate analyses of the five plausible values. A within-school sampling weight was used in this analysis to account for differential probability of sampling students. In brief, the calculation of school proficiency rate in NAEP considered the measurement error and sampling weight.

Once school proficiency rates were calculated from the NAEP and state assessment data-sets, an empirical search was conducted to estimate an ‘adequate’ level of school funding that corresponds to the target proficiency level in eight-grade mathematics. School districts’ current instructional PPE were extracted from the 1996 F-33 data. Current expenditures are expenditures for the day-to-day operation of schools and school districts. They include expenditures for instruction, support services, food services, and enterprise operations. In addition, they exclude expenditures for capital outlays (e.g. school construction, renovation, and equipment) and programs outside the regular pre-school to Grade 12 scope (e.g. adult education, community services, etc.). It should be noted that this study focuses on expenditures for instruction that includes expenditures for activities of teachers and instructional aides or assistants engaged in regular instruction, special education, and vocational education programs. It excludes any spending for other non-instructional or administrative purposes. In 1996, instructional expenditures comprised about 61% of current educational expenditures in an average Kentucky school district and about 67% in an average Maine school district.

Using the ‘successful schools’ approach for the first approximation (see Augenblick, Alexander, and Guthrie 1995; Augenblick, Myers, and Anderson 1997; Guthrie and Rothstein 1999), schools where at least 30% of students performed at or above the Proficient level separately on the NAEP and state assessments were identified. In the absence of state-imposed measurable performance target for schools prior to NCLB, 30% was chosen as the threshold of high performance because it was reasonably above the statewide norms (presumably within the reach of many schools’ capacity) at the time of testing.

Of these relatively high-performing schools, PPE were averaged only for those schools whose PPE measures were not extremely high or low for the state as a whole. Specifically, the analyses of school funding adequacy excluded extreme schools as outliers – that is, ones below the 10th percentile and above the 90th percentile for at least one of these three variables: PPE, median value of housing, and percentage of children below poverty line (see Appendix 1, Table A2). This selection was designed to calculate average expenditures from only high-performing schools with normal funding conditions; it may reduce the possibility of overfunding or underfunding by excluding districts that deviate considerably from the norms.

Despite the advantages of a straightforward and transparent estimation procedure, this empirical observation approach has some drawbacks. It does not explicitly take into account differences among schools in their student characteristics. It is assumed that, for example, schools with large proportions of economically disadvantaged and/or racial minority students can produce the same level of proficiency with the same resources as schools with smaller proportions of such students (Guthrie and Rothstein 1999). The same proficiency measure is applied to schools based on the presumption that, regardless of racial and socio-economic status, all schools must achieve the same average proficiency to be considered as having adequate outcomes.

Given these shortcomings, this study also used the cost function analysis method to provide information about the contributions of various characteristics of school
districts to the costs of education. In algebraic terms, educational cost function can be represented by the following equation (see Reschovsky and Imazeki 2004):

\[ E_i = f(S_i, P_i, Z_i, F_i, e_i, u_i), \]

where PPE \((E_i)\) are specified as a function of public school outputs \((S_i)\), a vector of input prices \((P_i)\), the characteristics of the student body such as percentage of students in poverty \((Z_i)\), other characteristics of the school district such as its size \((F_i)\), a vector of unobserved characteristics of the school district \((e_i)\), and a random error term \((u_i)\).

One of challenges for this line of research is to select performance measures that are rich enough to capture success in a range of educational activities and to give them appropriate weights for the construction of valid composite performance index (Duncombe and Yinger 1999). Previous studies used different mixes of school outputs with focus on reading and mathematics performance indicators. In their cost function analysis of data from New York State, Duncombe, Ruggiero, and Yinger (1996) constructed an educational performance index based on weighted average of three indicators, including reading and mathematics proficiency rates, Regents diploma rate, and graduation rate. In their cost function analysis of data from Texas, Reschovsky and Imazeki (1998) used the state assessment composite pass rate and average college entrance exam scores.

This study chose to focus on one subject area, mathematics. This decision was made primarily because of inherent limitation of the NAEP design that tested different samples for different subjects and thus did not allow for linking test results in multiple subjects at the school level. Nevertheless, there can be some rationales for using only mathematics data in this study: (1) mathematics is the core-subject upon which state education reform has focused during the 1990s for academic improvement at the middle level of education (Lee 1998); (2) mathematics achievement is influenced more directly by school effects than by family effects (Murnane 1975); and (3) mathematics has relatively stronger linkages between NAEP and state assessments with their common alignment with the national curriculum framework such as the National Council of Teachers of Mathematics (1989) standards.

Estimation of the cost function must take into account the fact that the educational output (mathematics achievement in this study) and PPE are determined simultaneously; local school board decisions to raise the level of student performance may affect the level of spending, while decisions concerning school funding are likely to influence performance (Reschovsky and Imazeki 1998, 2004). Given this reciprocal causation problem, two-stage least squares regression was used to estimate cost functions. This method treats the school average mathematics proficiency variable as endogenous, and uses instrumental variables that are assumed to be highly correlated with school proficiency but uncorrelated with theoretical errors of educational cost as the dependent variable. While selection of the best instruments for school performance is difficult, previous educational cost function studies drew on a set of variables that were related to the demand for public education (Duncomb, Ruggiero, and Yinger 1996; Reschovsky and Imazeki 2004). According to the median voter model, demand for public education is seen as a function of median household income and educational preferences. As proxies for educational preferences, school districts’ social and racial composition variables are used as instrumental variables: school average socio-economic status (SES) and percentage of White students in schools.\(^5\)
Instructional spending per pupil was specified as a function of predicted mathematics proficiency and other factors that directly influence the cost of education, including property value, poverty rate and district size. Property value represents a demand variable: the greater the local wealth, the greater local financial support for education. Controlling for property value, the percentage of students below the poverty line may mean a greater need for spending. The cost of raising the achievement of socio-economically disadvantaged students, who are at a greater risk of academic failure, is higher. The number of children and schools in the district also represent the variation in cost. Larger enrollment will lead to lower costs if there are economies of scale. Controlling for enrollment, more schools may lead to higher costs because of the fixed costs of running each school (Ferguson and Ladd 1996). Data on school district size, housing property value and poverty rate were drawn from the 1990 Census available in the 1996 Common Core of Data. Data on school input prices such as teacher salary were not available.

This study has several limitations. First, the study uses proficiency rates in a single subject (mathematics) as measures of school output. School cost functions generally assume there to be multiple outputs of schools, usually related to the variety of subjects. Thus, a cost function that only includes performance on one output will potentially provide misleading results. This study’s findings should be taken with caution for this reason. While reading and mathematics test scores remain key indicators of school performance under the NCLB, their relationships with each other or with measures of performance in other subjects need investigations. In the state assessment data used in this study, the correlation between school average mathematics and science performance indicators was 0.65 in Maine and 0.86 in Kentucky. The correlation between reading and mathematics performance indicators was 0.71 in Maine. Although there are generally positive interrelationships in school performance among multiple subjects, schools may have relative strengths or weaknesses in different areas. The sensitivity of cost estimation to different mixes of and weights to multiple performance indicators needs to be known.

Second, this study uses school average mathematics proficiency in a particular year (1996) as a measure of school output. Although this approach is in line with the current NCLB requirement, these measures of school performance represent ‘status’ as opposed to ‘change’, and thus they do not capture value-added contributions of schools to making progress in student achievement (see Reschovsky and Imazeki 2004). Under this value-added approach, only the part of test scores that is independent of these student/family characteristics may be considered a school output (Meyer 1997). However, the data collected under NAEP does not allow for tracking school performance annually. Further, the study’s use of school average SES and percentage of White students as instruments for proficiency rate also raises a question about their validity, since these background variables are well correlated with performance status but not so with gain.

Third, this study does not fully examine the issue of inefficiency as one potential source of expenditure variation. Apparently, how much schools spend on education is not synonymous with how much it costs schools to provide those services. Although the study attempted to exclude high spending districts that may produce adequate outcomes inefficiently, more sophisticated approach is needed to develop an index of efficiency and control for differences among districts with that index (see Duncombe, Ruggiero, and Yinger 1996; Ruggiero 2001).
This study also has several data-specific limitations, which need to be considered in interpreting its results. First, the study uses only part of the NAEP data for estimation of school performance proficiency, which provides information on mathematics achievement outcome for one particular grade. Second, there are also limitations in combining F-33 data with NAEP. F-33 data are only available at the district level, and within-district variation in school funding may be ignored. While few school systems collect budget information at the school level, there is a growing need for school-level spending data (see NCES 1996). Kentucky is one of pioneering states that currently collect and report school-level per-pupil spending in their school accountability report cards. Third, although schools that participate in NAEP are randomly selected and represent the school population in each state, using this overlap sample of schools common to both NAEP and state assessment can bring potential bias in estimating school funding adequacy.

Results

Comparison of the NAEP and state eight-grade mathematics assessment results in 1996 revealed how well students in Kentucky and Maine met national versus state performance standards. As shown in Figure 2, the percentage of students at or above the NAEP Proficient level is smaller than that of students at or above the MEA Advanced level: 31% of Maine eighth-grade students meet the NAEP’s Proficient level in mathematics as of 1996, whereas only 9% of the students meet the MEA’s Advanced level. Whether we base our judgment of Maine students’ performance on the NAEP or MEA achievement levels, we come to the same conclusion that a majority of the eighth-grade student population in Maine did not perform at the desired achievement level in mathematics. Comparison of 1996 NAEP and KIRIS assessment results also reveals inconsistent performance results in Kentucky. The percentage of eighth-grade Kentucky students at or above Proficient in mathematics is 28% on KIRIS and 16% on NAEP (see Figure 2). In brief, the performance standard for the MEA was set at a relatively higher level than the standard for NAEP, whereas the performance standard for the KIRIS was set at a relatively lower level than the standard for NAEP.

Since the data were collected in 1996, there were some changes to both states’ assessments. For example, Kentucky changed its state assessment from KIRIS to the Commonwealth Accountability Testing System (CATS) in 1999; the CATS includes multiple tests, among which the Kentucky Core Content Test – a criterion-referenced test – is the successor of KIRIS (see Lee 2007). While the performance standard remains higher on the NAEP than on the state assessment, the gap narrowed slightly between 1996 and 2003. In 2003, the percentages of Grade Eight students ‘proficient’ in mathematics were 31 on the CATS, as compared with 24 on NAEP. Maine changed its performance standards for MEA in 1999, but the state still remains to have relatively higher performance standards for its own state assessment than the NAEP counterpart. In 2003, the percentages of Proficient students (meeting or exceeding the standards in Maine’s Learning Results) in Grade Eight mathematics were 18 on MEA and 29 on NAEP.

Notwithstanding these statewide aggregate patterns of the discrepancies in Kentucky and Maine, there are substantial variations among schools in terms of the direction and/or size of the gap between NAEP-based proficiency rate and state assessment-based proficiency rate in 1996 eighth-grade mathematics (see Figure 3).
While the direction of the NAEP versus state assessment gap clearly diverged between the two states (mean = −11.91 in Kentucky; mean = 20.66 in Maine), both states showed a similar degree of variation among schools in the size of the gap (standard deviation = 12.26 in Kentucky; standard deviation = 11.32 in Maine). The distribution of the school-level proficiency gap tended to be positively skewed in Kentucky but negatively skewed in Maine.

It was hypothesized that the discrepancies between NAEP and state assessments in terms of school proficiency would result in differential estimation of school funding adequacy. Table 1 presents the estimates of base instructional costs in 1996 dollars for
Kentucky and Maine by the type of assessment (row) and the estimation method (column) used. The average per-pupil instructional expenditure of high-performing schools in Kentucky based on NAEP was $4042 ($5254 in 2005 dollars).7 When we applied the same criterion of 30% proficiency for selection of high-performing Kentucky schools based on KIRIS, the estimate of adequate PPE would decrease to $3910 ($5083 in 2005 dollars). The NAEP-based estimate of adequacy level exceeds its counterpart based on the state assessment by $132. The average per-pupil instructional expenditures of high-performing schools in Maine based on NAEP were $4601 in 1996 dollars ($5981 in 2005 dollars). When we applied the same criterion of 30% proficiency for selection of high-performing Maine schools based on MEA, the estimate of adequate PPE would increase to $5007 ($6509 in 2005 dollars). The difference between the NAEP-based adequate base funding level and its counterpart based on the state assessment in Maine was $406.

Further estimation of instructional costs was conducted through a two-stage least-squares regression analyses. Table 2 summarizes the results of the first-stage multiple regression using schools’ racial and socio-economic composition of student body and district characteristics to predict individual schools’ eighth-grade mathematics

<table>
<thead>
<tr>
<th></th>
<th>‘Successful Schools’ method</th>
<th>‘Cost Function’ method</th>
<th>Gap between methods</th>
</tr>
</thead>
<tbody>
<tr>
<td>Maine</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>NAEP-based PPE</td>
<td>4601</td>
<td>4734</td>
<td>133</td>
</tr>
<tr>
<td>MEA-based PPE</td>
<td>5007</td>
<td>5421</td>
<td>414</td>
</tr>
<tr>
<td>Gap between NAEP and MEA</td>
<td>406</td>
<td>687</td>
<td></td>
</tr>
<tr>
<td>Kentucky</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>NAEP-based PPE</td>
<td>4042</td>
<td>4269</td>
<td>227</td>
</tr>
<tr>
<td>KIRIS-based PPE</td>
<td>3910</td>
<td>4049</td>
<td>139</td>
</tr>
<tr>
<td>Gap between NAEP and KIRIS</td>
<td>132</td>
<td>220</td>
<td></td>
</tr>
</tbody>
</table>

Table 2. First-stage regression results for cost function analysis: 1996 NAEP and state assessment (KIRIS and MEA) eighth-grade mathematics proficiency rates as a function of schools’ racial and socio-economic compositions and district characteristics in Kentucky and Maine.

<table>
<thead>
<tr>
<th>Kentucky</th>
<th>Maine</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>KIRIS-based Proficiency</td>
</tr>
<tr>
<td>Intercept</td>
<td>−5.49</td>
</tr>
<tr>
<td>Proportion of white students in school</td>
<td>30.85*</td>
</tr>
<tr>
<td>School average socio-economic status</td>
<td>8.33*</td>
</tr>
<tr>
<td>Total number of schools in district</td>
<td>−1.78</td>
</tr>
<tr>
<td>Total number of children in district</td>
<td>0.002</td>
</tr>
<tr>
<td>Median value housing</td>
<td>0.000</td>
</tr>
<tr>
<td>Percentage of children below poverty</td>
<td>0.17</td>
</tr>
<tr>
<td>$R^2$</td>
<td>0.36</td>
</tr>
</tbody>
</table>

Note. *p < 0.05, **p < 0.01, ***p < 0.001.
proficiency rates based on state assessment versus NAEP. Building on the results of the first-stage regression, the second-stage regression was run by using predicted values of school eighth-grade mathematics proficiency (one estimated from state assessment and another estimated from NAEP) to explain instructional expenditures per pupil as dependent variable. Table 3 summarizes the results of the second-stage multiple regression using those predicted values of proficiency for estimation of the cost functions in Kentucky and Maine. From the cost functions derived from state assessment-based proficiency data and from NAEP-based proficiency data, adequate instructional spending levels were estimated for an average high-performing school that would possess the state average characteristics of school funding conditions (district size, property value, and poverty level) and at the same time achieve a 30% or higher level of proficiency rate.

As shown in Table 1, there was very little difference between the two estimates of instructional expenditures per pupil deemed adequate in Kentucky; that is, $4049 ($5264 in 2005 dollars) based on the KIRIS proficiency standard versus $4269 ($5550 in 2005 dollars) based on the NAEP proficiency standard. Similarly, Table 1 summarizes the estimates of adequate instructional expenditures per pupil in Maine as derived from the state’s cost function reported in Table 3. The school in Maine with average costs (i.e. where each of the cost factors is set equal to its mean) must spend $4734 ($6154 in 2005 dollars) per pupil for instruction to reach the standard of 30% proficiency on NAEP. The figure is raised to $5421 ($7047 in 2005 dollars) to reach the same performance standard (30% proficiency level) on MEA.

Discussion

There have been policy movements toward performance-based school accountability and funding across the nation and states during the past decade. However, there remains a gap between policy goals and technical realities. Despite the imperative of using multiple measures for school accountability, it was not clear what measures should be included and how that information should be combined into an overall

### Table 3. Second-stage regression results for cost function analysis: 1996 PPE as a function of predicted eighth-grade mathematics school proficiency and district characteristics in Kentucky and Maine.

<table>
<thead>
<tr>
<th></th>
<th>Kentucky PPE</th>
<th>Maine PPE</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>KIRIS-based</td>
<td>NAEP-based</td>
</tr>
<tr>
<td></td>
<td>Proficiency</td>
<td>Proficiency</td>
</tr>
<tr>
<td></td>
<td>as a predictor</td>
<td>as a predictor</td>
</tr>
<tr>
<td>Intercept</td>
<td>2512.26***</td>
<td>2480.94***</td>
</tr>
<tr>
<td>Total number of schools in district</td>
<td>-44.36*</td>
<td>-49.81**</td>
</tr>
<tr>
<td>Total number of children in district</td>
<td>0.05*</td>
<td>0.06**</td>
</tr>
<tr>
<td>Median value housing</td>
<td>0.01**</td>
<td>0.01**</td>
</tr>
<tr>
<td>Percentage of children below poverty</td>
<td>28.54***</td>
<td>30.71***</td>
</tr>
<tr>
<td>Predicted eighth-grade mathematics proficiency</td>
<td>2.31</td>
<td>3.01</td>
</tr>
<tr>
<td>$R^2$</td>
<td>0.83</td>
<td>0.83</td>
</tr>
</tbody>
</table>

Note. †$p < 0.10$; *$p < 0.05$; **$p < 0.01$; ***$p < 0.001$. 
evaluation of school performance. Moreover, despite the appeal of tying school performance to funding, it was not clear what the implications are of possible discrepancies among those multiple measures for determination of adequate school funding. The discussion of measurement issues has not yet informed school finance reform policy debate. The past studies of school funding adequacy all relied on the current state performance standards and took for granted the states’ own assessment as a tool to measure school performance. This suggests that the educational research community work together to bridge the gap between the fields of educational testing and finance.

The purpose of this study was not to obtain accurate estimates of adequate school funding, but rather to unravel discrepancies between NAEP and state assessment results that can lead to quite different measures of funding adequacy. This study does not investigate causes of the discrepancies between NAEP and state assessments, but instead focuses on examining the consequences of these discrepancies for funding issues. Another study that examined possible contributing factors with the same states’ data pointed out differences in the definitions of performance standards and the methods of standard setting (Lee 2007). Further study is needed to examine the mechanism of what cause discrepancies between the national and state standards, and how they produce gaps in academic proficiency which in turn leads to gaps in funding adequacy.

The study’s estimates of instructional cost per pupil for meeting the target proficiency rate of 30% in Kentucky and Maine were approximately in the range of $5000–7000 (in 2005 dollars). The contrast of the two states that shared a common assessment framework, but differed in terms of educational accountability systems and policies, shows that the estimation of adequate funding is affected by not only the size of discrepancies between NAEP and state assessment results on student proficiency, but also the strength of the relationship between school performance and funding. The MEA-based estimate of adequacy was higher than the NAEP-based measure in Maine, where the state performance standard was higher than the NAEP performance standard for eight-grade mathematics in 1996. In contrast, the KIRIS-based estimate of adequacy was smaller than the NAEP-based adequacy in Kentucky where the state performance standard was lower than the NAEP performance standard for eighth-grade mathematics in 1996.

Higher standards will obviously require more funding. Therefore, the above results are predictable and should come as no surprise to policy-makers. However, the study reveals interesting patterns of differences between the two states. Why does Maine have a larger difference between NAEP-based and state assessment-based school funding adequacy while Kentucky has very little discrepancy? This may be attributable to two factors. One factor is related to the fact that the discrepancy between NAEP and state assessment results on proficiency itself was larger in Maine and in Kentucky. At the same time, another contributing factor is related to the fact that the relationship between PPE and achievement was significant in Maine but not in Kentucky.

The choice of assessment with different level of performance standard (NAEP versus state assessment in this case) would only matter when school performance is systematically associated with educational expenditures. In other words, different performance standards could result in different estimates of adequate school funding to the extent which the level of funding is closely linked to the level of performance. It is revealed by the finding that the difference between NAEP-based and state
assessment-based adequacy levels was larger in Maine than in Kentucky. While the estimation of adequate school funding was congruent with hypothesized direction in both states, the difference in estimated instructional cost by the type of assessment was significant in Maine but not in Kentucky.

There are two plausible explanations for the lack of significant relationship in Kentucky. First, schools in Kentucky may not have spent money effectively and thus money did not matter much for proficiency. This challenges the expectation that high-stakes testing would drive schools to spend money more effectively given greater incentives for raising student achievement. Second, the earlier statewide school funding reform in Kentucky may have equalized school expenditures so that the reduction of variation among schools in spending attenuated the relationship between PPE and achievement. However, this explanation does not appear to be supported by the data since there were very small differences in the measures of school funding equity between the two states as of 1996: wealth neutrality score was 0.072 in Kentucky and 0.064 in Maine, while the coefficient of variation was 9.8% in Kentucky and 13.5% in Maine (Quality Counts 2000).

The study also suggests that using different analytical methods can result in different estimates of adequate school funding. The choice of empirical observation of successful schools versus econometric analysis of cost function method appears to produce inconsistent cost estimates. The difference of estimation is caused by the ways in which the two methods identify adequate PPE. The successful schools method describes what the typical high-performing districts spend and it includes all schools that meet a certain performance criterion regardless of their other conditions. In contrast, the cost function method estimates the minimum cost of producing adequate performance outcomes, and it identifies an average school that meets desired performance standard under normal conditions. The difference due to this methodological choice effect can be as large as the difference due to the above-mentioned assessment choice effect.

The findings of this study demand that educational policy-makers become fully aware of problems with ‘racing to the bottom’ among states under dual standards of school performance, which threaten successful implementation of the NCLB goals. There is a dilemma in choosing one measure over the other for school accountability. On one hand, using only state assessments with relatively lower standards as the basis of school accountability decisions runs the risk of generating false impressions of academic proficiency and driving underfunding for schools. It will ultimately lead to failure to meet the desired national performance standards. On the other hand, simply raising state performance standards without addressing current school funding gaps will increase the number of schools that fail to meet desired performance targets. If states choose to tighten the alignment of state performance standards with national standards, standards for school funding adequacy also need to be adjusted accordingly in order to avoid underfunding.

Further, it is crucial for states to implement more evidence-based school improvement strategies and strengthen the connection between school resources and performance so that the adequacy of school funding can be established more firmly. Any school funding increases necessary to meet the standards can be offset by reductions in inefficiency due to wasteful spending. It is important to acknowledge that there can be areas of inefficiencies in current educational spending; for this reason, the future adequacy study should address not only how much high-performing districts or schools spend but also how efficiently they spend to accomplish a desired proficiency level.
Finally, the study calls for subsequent studies using more recent NAEP and F-33 data to investigate changes in school performance and funding relative to new standards under NCLB. The results might be sensitive to the timing of data collection, as performance standards adopted for school accountability system by states are likely to become moving targets. While states are permitted to shape schedules of improvement flexibly under NCLB, many states including Kentucky and Maine adopted a stair-step approach by choosing to keep AYP targets unchanged or to raise them slowly in the first phase. Subsequent studies need to update the estimates of instructional costs that can increase more rapidly as the states’ AYP targets move closer toward the goal of 100% proficiency by 2014.

Notes
1. Reschovsky and Imazeki (2001, p. 378) discuss the study by Augenblick, Myers, and Anderson (1997) that excluded all districts with particularly high and low levels of property wealth and of per-pupil spending: ‘Although this method provides a norm level of per-pupil spending, no systematic attempt is made to measure the likely variation in the costs of adequacy due to characteristics of individual school districts and their students.’ Verstegen (2002, p. 778) also points out several potential problems with this type of study: ‘If the funding system is inadequate for most districts in the state, as was the case in Kentucky, then a study that attempts to correlate current spending in select districts to student outcomes will generally fail to capture actual costs. Moreover, what a district spends and what resources/standards may cost can be separate issues that are muddled in these types of analyses. Still, another difficulty with this approach is that it defines efficient district as neither high nor low spenders when outliers are removed, leading to the possibility of recommendations that underfund education.’

2. Category labels and brief generic definitions of performance standard for each assessment are as follows:
   - **NAEP Proficient**: students demonstrate competency over challenging subject matter and are well prepared for the next level of schooling.
   - **KIRIS Proficient**: the student understands the major concepts, can do almost all of the task, and can communicate concepts clearly.
   - **MEA Advanced**: Maine students successfully apply a wealth of knowledge and skills to independently develop new understanding and solutions to problems and tasks.

3. The F-33 survey is part of the US Census Bureau’s ‘Survey of Local Governments.’ It is a universal survey of school funding in the nation and states (see O’Leary and Moskowitz 1995). The district level is the lowest level of aggregation in the F-33 data, and no nationally representative database exists that measures different types of expenditures at a lower level of aggregation. Because the school funding measure from the F-33 data is available only at the district level, the measure of PPE would simply vary among school districts within each state but not within districts (among schools and classrooms).

4. In fact, the study’s 30% proficiency rate as the minimum school performance target was higher than the states’ current annual measurable objective in eighth-grade mathematics to meet NCLB as of the 2004/05 school year: 26.93% in Kentucky (Kentucky Department of Education 2004) and 13% in Maine (Maine Department of Education 2003). In order to reach the ultimate target of 100% proficiency by the 2013/14 school year, both states plan to take a stair-step approach that will allow schools to improve slowly in the first few years as improvements are phased in, and then more dramatically once improvements are fully implemented.

5. SES is a standardized measure of student’s family and school’s SES level, and it is derived from the NAEP data. It is a factor composite of parents’ education (PARED), availability of reading materials at home (HOMEEN2), and school median income (MEDINC). Student-level factor loadings are as follows: PARED, 0.77; HOMEEN2, 0.68; MEDINC, 0.63. The factor has an eigenvalue of 1.46 and explains 49% of the combined variance. The underlying assumption is that these variables are related to voters’ willingness to pay for
education and only indirectly affect the cost of education through their effects on school performance (eighth-grade mathematics). Hausman test of endogeneity rejects the null hypothesis that there is no simultaneity regarding the relationship between instructional cost and school performance. This supports the use of instrumental variable estimation method.

6. Obviously disadvantaged groups of students are more expensive to educate but currently there are no national standards for determining the cost of educating different student groups. One common practice in school funding formula involves adjusting fall student enrollment by weighting poor students by 1.4, special education students by 2.3, and limited English proficiency students by 1.2.

7. The Employment Cost Index of the US Bureau of Labor Statistics was used to convert 1996 dollars into 2005 dollars. The Employment Cost Index for ‘Total compensation: Elementary and secondary schools: State and local government’ has increased 1.3 times from 1996 to 2005. It needs to be noted that the study’s estimates of instructional costs are not directly comparable with those from previous studies of school funding adequacy in Kentucky and Maine because they used different criteria, measures, and methods. Previous studies used evidence-based or professional judgment methods to estimate base costs of education (Quality Counts 2005). A study that used an evidence-based method estimated an average of $6893 per pupil as the basic cost in Kentucky to reach the state’s goal of having all students perform at or above ‘proficient’ on the state assessment by 2014 (Odden, Fermanich, and Picus 2003). A study of funding adequacy in Maine that used a professional judgment method estimated the per-pupil operating cost of $4543 for middle-school level regular students (Silvernail 2000).

References
Fuller, B., K. Gesicki, E. Kang, and J. Wright. 2006. Is the No Child Left Behind Act working?: The reliability of how states track achievement. PACE Working paper 06-1, University of California, Berkeley.


### Appendix 1.

Table A1. Distributions of Maine and Kentucky public school samples in NAEP 1996 8th grade math by sampling strata.

<table>
<thead>
<tr>
<th>District size</th>
<th>School size</th>
<th>Urbanization</th>
<th>Percentage of minority</th>
<th>Number of originally selected schools</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Kentucky</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Small</td>
<td>Small</td>
<td>Rural</td>
<td>None</td>
<td>1</td>
</tr>
<tr>
<td>Small</td>
<td>Large</td>
<td>Mid-size city</td>
<td>Low</td>
<td>6</td>
</tr>
<tr>
<td>Small</td>
<td>Large</td>
<td>Mid-size city</td>
<td>Medium</td>
<td>6</td>
</tr>
<tr>
<td>Small</td>
<td>Large</td>
<td>Mid-size city</td>
<td>High</td>
<td>5</td>
</tr>
<tr>
<td>Small</td>
<td>Large</td>
<td>Urban fringe</td>
<td>Low</td>
<td>6</td>
</tr>
<tr>
<td>Small</td>
<td>Large</td>
<td>Urban fringe</td>
<td>Medium</td>
<td>5</td>
</tr>
<tr>
<td>Small</td>
<td>Large</td>
<td>Urban fringe</td>
<td>High</td>
<td>5</td>
</tr>
<tr>
<td>Small</td>
<td>Large</td>
<td>Large/small town</td>
<td>None</td>
<td>43</td>
</tr>
<tr>
<td>Small</td>
<td>Large</td>
<td>Rural</td>
<td>None</td>
<td>33</td>
</tr>
<tr>
<td><strong>Maine</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Small</td>
<td>Small</td>
<td>Mid-size city/urban fringe</td>
<td>None</td>
<td>1</td>
</tr>
<tr>
<td>Small</td>
<td>Small</td>
<td>Small town</td>
<td>None</td>
<td>2</td>
</tr>
<tr>
<td>Small</td>
<td>Small</td>
<td>Rural</td>
<td>None</td>
<td>13</td>
</tr>
<tr>
<td>Small</td>
<td>Large</td>
<td>Mid-size city/urban fringe</td>
<td>None</td>
<td>16</td>
</tr>
<tr>
<td>Small</td>
<td>Large</td>
<td>Small town</td>
<td>None</td>
<td>53</td>
</tr>
<tr>
<td>Small</td>
<td>Large</td>
<td>Rural</td>
<td>None</td>
<td>33</td>
</tr>
</tbody>
</table>


Table A2. Descriptive statistics of variables used for the analysis of school mathematics proficiency and funding adequacy in Kentucky and Maine.

<table>
<thead>
<tr>
<th></th>
<th>Kentucky</th>
<th></th>
<th>Maine</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$n$</td>
<td>Mean</td>
<td>Standard deviation</td>
<td>$n$</td>
</tr>
<tr>
<td>Proportion of white students$^1$</td>
<td>90</td>
<td>0.86</td>
<td>0.16</td>
<td>92</td>
</tr>
<tr>
<td>Average SES$^1$</td>
<td>90</td>
<td>0.45</td>
<td>0.53</td>
<td>92</td>
</tr>
<tr>
<td>Current instructional expenditures per pupil$^2$</td>
<td>90</td>
<td>3853.11</td>
<td>425.37</td>
<td>88</td>
</tr>
<tr>
<td>Total number of schools in district$^3$</td>
<td>90</td>
<td>35.31</td>
<td>50.43</td>
<td>92</td>
</tr>
<tr>
<td>Total number of children in district$^3$</td>
<td>69</td>
<td>36,946.15</td>
<td>52,143.57</td>
<td>92</td>
</tr>
<tr>
<td>Median value housing$^3$</td>
<td>69</td>
<td>49,797.52</td>
<td>13,988.78</td>
<td>92</td>
</tr>
<tr>
<td>Percentage of children below poverty$^3$</td>
<td>69</td>
<td>24.14</td>
<td>9.84</td>
<td>92</td>
</tr>
<tr>
<td>Proficiency rate on NAEP assessment$^1$</td>
<td>90</td>
<td>14.86</td>
<td>11.86</td>
<td>92</td>
</tr>
<tr>
<td>Proficiency rate on state assessment$^4$</td>
<td>90</td>
<td>26.77</td>
<td>12.92</td>
<td>92</td>
</tr>
</tbody>
</table>

Note: Multiple sources of data were used to obtain or derive the above variables: $^1$1996 NAEP eighth-grade mathematics assessment data; $^2$1996 F-33 data; $^3$1996 Common Core of Data (1990 census); $^4$1996 KIRIS or MEA eighth-grade mathematics assessment data.