

The Cost of Accountability: Considerations and Assumptions

Yas Nakib, Ph.D.

George Washington University

Patrice Iatarola, Ph.D.

Florida State University

Introduction

How expensive are education accountability systems? Where does the money come from? Where does it go? Could the same resources be better spent elsewhere in the system? These questions have been raised with increasing frequency as states and districts expand their testing systems by adding more grades, more students, more subjects and more kinds of tests. Estimating the cost of accountability and assessment in Florida and elsewhere is complex. The direct cost of accountability could be minimal compared to overall costs, as several analyses have demonstrated. But any assessment of direct costs is likely to be incomplete; indirect costs as well as the cost of opportunities not pursued could be much higher. Finally, costs should not be viewed alone but in relation to benefits.

The purpose of this brief is to examine these complexities and outline a conceptual framework for conducting a sound estimation within the context of Florida's accountability system. The brief addresses some of the possible costs of not having accountability systems. It focuses on Florida's accountability system and its related assessments. It does not place a dollar figure on the system; rather it lays out the system components, their relation to each other and discusses the critical role that a few key assumptions about accountability can play in determining costs.

Assessment in Florida: Sunshine State Standards and FCAT

The Florida Comprehensive Assessment Test (FCAT) is part of Florida's effort to tie teaching and learning to educational standards. It was designed to assess student achievement of high-order cognitive skills represented in the Sunshine State Standards (SSS) in reading, writing, mathematics, and science. It has two purposes. The first is to assess student mastery of the SSS using a criterion-referenced test. A criterion-referenced test measures a student's performance relative to expected levels of achievement; a norm-referenced test measures a student's performance relative to the performance of other students. The second purpose is to compare the performance of Florida students to the reading and mathematics performance of students across the nation using a norm-referenced test (FCAT-NRT, a form of the Stanford 9). FCAT reading and mathematics is administered to students in grades 3-10. FCAT writing and science is given in only 3 grades, one each in elementary, middle and high school.

Cost Estimates of Accountability Systems

Many studies and estimates have examined the cost of testing in states across the nation. Some have reported a very low cost relative to average per-pupil expenditure (Hoxby, 2002; Phelps, 2000). In an oft-cited study, Hoxby analyzed the educational accountability costs of 25 states that implemented fairly rigorous systems of accountability by 2001. Adjusting for factors such as the specific testing requirements in the state and the size of the state's population, Hoxby's estimates of the costs of accountability systems ranged from a low of \$1.79 per pupil (South Carolina) to a high of \$34.02 per pupil (Delaware). Hoxby asserts that even if every state spent as much per pupil on accountability as Delaware, the cost would still amount to less than one half percent (<0.5%) of the total per-pupil expenditure in the nation's public schools. She compares the costs of accountability to those of class size reduction to argue that the costs of accountability, relative to its benefits, are much lower than for other school reforms.

Cost Generating Components of Accountability Systems

Accountability is a generic term referring to multiple policies and activities. Decisions as to what components of the system should be included in any estimation of accountability costs have serious implications for cost estimates. Testing is clearly a major component with discrete cost implications, but "accountability" also includes initiatives such as standards-based reform and decentralized decision making. In Florida, it also includes school choice, merit-based financial incentives, increased retention, enhanced services for low-performing students and professional development.

Table 1 describes the three major elements to be considered in estimating the cost of testing and assessment under the Florida *A+ Plan*: (1) costs, (2) benefits, and (3) assumptions. As described in the sections that follow, costs are not always easy to measure and should be estimated only after clearly stating the assumptions on which they are based. Once estimated, costs should always be evaluated relative to the benefits produced.

Table 1

Consideration	Examples
I. Costs	
- Direct costs	<ul style="list-style-type: none"> Test-related activities: Test development, training, test administration, scoring, reporting, information Elements of the <i>A+ Plan</i> and NCLB triggered by test results <ul style="list-style-type: none"> - School grading (A-F) - 3rd grade retention based on FCAT reading performance - School choice provisions for students at low-performing schools
- Indirect costs	<ul style="list-style-type: none"> - Narrowing of curriculum to the test skills and components. - Test "fatigue" by students, teachers and others. - Parental time
- Opportunity costs	- Benefits foregone by using resources on assessment rather than activities, such as instruction
- Level of costs	National, state, local (district, school, classroom), private
- Kinds of costs	Personnel (teacher, administrator, other professionals), materials, supplies, time.
- Infrastructure, capacity-building, and buy-in	Resources, training and technology required to support testing and assessment; activities required to get principals, teachers and district officials to accept, implement, and use assessments
- Time period	Initial versus lifetime costs
- Testing format	Multiple choice, written responses, other
- Economies of scale	Number, type, grades of students tested; frequency of the test
- Cost variations	Cost variations among districts and schools depending on size, capacity, student population, teacher characteristics/turnover, student mobility
II. Assumptions	
	For example: <ul style="list-style-type: none"> - Tests are sufficiently standardized and administered in the same way to all students - Reliability and validity of tests are regularly monitored and checked. - Tests cause minimal disruption of regular activities - The marginal cost of test development and administration is minimal
III. Benefits	
	<ul style="list-style-type: none"> - Provide screening, diagnosis and progress monitoring data to guide instruction and curriculum decisions - Provide information for targeting interventions and resources to low-performing schools - Assess the mastery of desired competencies - Foster accountability on the part of students, parents, teacher and administrators

Frameworks for Measuring the Cost of Assessment

Picus (1994) presents one of the most comprehensive frameworks for analyzing the cost of assessment. He considers costs using a matrix of three dimensions: 1) components of the assessment, 2) level at which expenditures occur, and 3) kinds (e.g., personnel, materials, and supplies) of items. Components include activities such as development, production, training, instruction, test administration, management, scoring, reporting, and program evaluation of the assessment. The level of expenditure can be national, state, district, school, classroom, or the private market. For each component and at each level,

different kinds of costs—e.g., personnel, materials, and supplies—are incurred. Phelps (2000) provides an alternative, but similar, framework. He asserts that when accounting for the cost of student testing activities, there are four major “objects” on which resources are expended: 1) purchased material and services, 2) time spent on the activities by schools, districts, and state agencies, 3) time of students preparing and taking the test, and 4) administrative and building overhead. Alternatively, one may account for the cost of testing according to the functions for which resources are used: start-up test development, ongoing test development, preparing students, training and preparing for administration, time spent taking or administering the test, training for scoring, scoring, collecting, sorting and mailing the tests, analyzing and reporting the scores, and miscellaneous other activities.

Estimating Costs: Direct and Related Costs

One of the reasons it is so difficult to estimate costs is that the assessment systems do not stand alone. First, costs are obscured in an existing comprehensive system of education, rendering it difficult to identify costs unique to assessment. Two, the system has multiple layers within the intergovernmental system, and all three levels—federal, state and local districts—have interlocking requirements for assessment. Three, the assessment system, by design, triggers other activities and these activities generate their own expenditures.

Listed below are conditions and concerns that must be addressed in estimating costs: **Test-related Activities.** Major test-related activities include test development, administration, scoring, and reporting. Phelps (2000) claims that the test-related costs most difficult to measure are personnel costs and the time devoted to test-related activities.

Development versus Implementation Costs. Because expenditures for test development are small relative to non-development expenditures (test administration, scoring, and reporting), estimates of expenditures may be lower in the first few years of test development than in later years. This is an important factor to take into account, because cost analyses should cover total costs over the lifetime of an initiative. The U.S. General Accounting Office (2003) finds student assessment costs increase in subsequent years as states begin to administer and score the tests and report the results. For example, in Florida, the state-level assessment cost per K-12 student grew from \$4.44 in 1996 to \$16.57 in 2004 (Florida Department of Education, 2004). Costs will also vary by the type of test instrument used; for example, scoring multiple-choice instruments is less costly than scoring instruments that include open-ended items.

State versus Local Costs. State-level costs of testing and assessment are easier to measure than local costs. Over a nine-year period (1996 to 2004), the state spent a total of \$227.7 million in state and federal funds on developing, administering, scoring, and reporting the FCAT (Florida Department of Education, 2004). The bulk of these expenditures were directed toward scoring the test and reporting the results. Those expenditures, however, represent only the state's basic expenditures to administer the test. They do not reflect all of the state costs (for example, they do not include training costs) nor costs that local districts and schools bear (for example, they exclude the cost of lost instructional time for testing and test preparation; the costs of distributing, securing, collecting, checking, shipping test materials; and the cost of distributing results to parents).

State versus Federal Costs. Another concern is how the Florida accountability system intersects with existing and new Federal accountability requirements, especially those stipulated by the *No Child Left Behind* (NCLB) legislation. In education, federal and state governments often have overlapping and sometimes contradictory roles and functions. In accounting for some of the costs imposed by the legislation's mandates, it could be quite difficult to distinguish between the state versus the federal government's responsibilities, including federal and state oversight.

Direct versus Indirect Costs. Estimates should not be limited to direct costs, such as student and teacher time devoted to testing and test preparation. Indirect costs, such as those listed in Table 1, may be significant and must be accounted for as well. These costs, however, may be difficult to measure in terms of both time and monetary value.

Direct and Indirect Costs versus Opportunity Costs. Estimating direct and indirect costs is quite challenging, but there are well-established cost-accounting principles that can be applied to allocate expenditures and time to the tasks of an accountability system. To estimate the true economic costs, however, one must estimate the foregone benefits of using the resources for another activity—the opportunity costs. For example, opportunity costs are borne at the local level by shifting resources from other established practices to assessment. An often-cited example is the cost of time which could be spent on instruction that teachers and administrators instead spend preparing for and administering the tests. Haney, Madaus, and Lyons (1993) estimate the opportunity cost of testing by using net present value of wages foregone as a result of test preparation and teacher and student time. It would be quite difficult to accurately account for such costs since many are dependent on student, school and district conditions that vary considerably across the state. A full accounting of opportunity costs must include estimates of the

benefits that would accrue to using the resources for one activity and not another—what benefits are foregone by using resources on accountability rather than other efforts?

Cost Implications of Notable Activities Triggered by Assessment Results

Elements of the *A+ Plan* Related to Assessment. The *A+ Plan* calls for grading of schools based on student performance on statewide assessments and, since 2002-03, testing in more subjects (science as well as reading, mathematics, and writing) and in more grades (3 through 10). The cost of monitoring and administering opportunity scholarships (vouchers) to students at persistently failing schools also should be added to the cost of accountability.

Student Retention. Since 2002-03, Florida law has required, with some exemptions, that third graders at the lowest level of proficiency in reading (i.e., scoring at level 1) be retained and given intensive reading instruction. The cost of holding back students may be significant: 23% of third-grade students who took the Florida assessment test scored at level 1 in reading in 2002-03; 14% of third graders that year were ultimately retained (C. Cross, personal communication, August 31, 2004). By not progressing through the system, retained students add to the number of students to be educated and to the length of time enrolled in public schools. Because these students will be re-tested, the marginal costs of testing—the costs associated with serving one more student—will at least double.

Student Mobility. Another confounding factor in efforts to estimate costs is student mobility within and across districts (Figlio, 2002). Florida already experiences high levels of student mobility. Both NCLB and the *A+ Plan* allow students in persistently low-performing schools to transfer to other schools or schools in other districts. Further, districts may modify their school zoning plans as a way of responding to accountability pressures resulting in yet more movement. Student mobility generates administrative costs for school systems as records and space must be accommodated in the receiving and losing schools and educational costs for students as they attempt to accommodate changing curricula, teachers and instructional strategies across the schools.

Sources of Cost Variation

Much of the actual cost of accountability is borne by entities other than state education agencies, such as districts and schools (Monk, 1994). At the district and school levels, costs may vary depending on existing capacities, student population, educators'

qualifications and need, student mobility, educator turnover, school and district size, and other conditions. The indirect costs of the foregone time of teachers and administrators, for example, is hard to estimate because costs vary among teachers with differing abilities who are teaching diverse student populations in varying grade levels.

Specific practices, such as teaching to the test, may also vary across districts, schools and even classrooms. To the extent that testing changes the behavior of teachers, the costs will vary depending on specific teacher characteristics, the school and district environment and other factors. For example, Figlio and Getzler (2002) found that some schools in Florida respond to accountability pressures by “gaming” the FCAT testing system by reshaping the test pool. With the introduction of the FCAT in 1996, schools in six large Florida districts tended to reclassify students (especially low-income and previously low-performing students) as disabled, thereby making them ineligible to contribute to the school's grade. Moreover, to the extent that high-stakes testing contributes to the decision by some students to drop out, this would further add to the indirect costs of accountability.

Costs are not necessarily borne the same way in classrooms, schools or districts within the same state. For example, estimates of expenditures on piloting the implementation of testing in one school district may not provide accurate estimates of a typical district's cost. These variations may emanate from any number of sources, including differences in the types of students, available resources, the intensity of implementation, the use of resources, and cost of living differences. Moreover, variations may stem not only from differences across districts, but also from differences across schools. Cost variations across schools due to student needs, preparation, remediation, resource availability, and many other factors should be carefully considered in estimating the total costs. The level of schooling should also be considered, since cost components differ in high schools relative to elementary schools. In addition, costs may vary significantly because of the size of a district, the composition of its schools, and the scope of the reform being implemented. For example, all else being equal, when fixed and variable costs of accountability borne by districts are distributed over a smaller number of students, the average and marginal costs tend to be higher.

Cost Assumptions

In estimating the cost of student assessment, it is important to clearly state the assumptions on which costs are based. Examples include:

1. The use of time is uniform across and within districts.
2. The tests are sufficiently standardized and administered in the same way to all students. They are uniform in their modes and process of assessment.

3. Reliability and validity of the tests are regularly monitored and checked.
4. The tests utilize some existing resources to administer.
5. The average test cost fairly represents the cost across students taking it.
6. The costs can easily be annualized, and apportioned in a way that would allow for attributing the costs to the various cost categories.
7. The tests cause minimal disruption of regular activities. The tests are administered during the regular school year, using regular school personnel, and are integral parts of the school system curricular and instructional plan.
8. Variation of costs borne by districts due to size and scope is minimal. The tests are used in school districts to replace, rather than supplement, any or most preexisting tests.

The validity of these and other assumptions should be carefully considered before proceeding with cost estimates. This task is not always easy, because the same set of assumptions may not be valid at all times. Nevertheless, cost estimates may vary widely depending on the assumptions made, not only about testing and assessment but also about the broader range of policies and activities subsumed under “accountability.”

Cost in Relation to Benefits

Costs alone will not provide adequate justification for the continuance or end of accountability systems. Costs must be viewed relative to the benefits generated by a program. To the extent that a program provides sufficient private and public benefits, costs may be considered worthwhile.

Advocates of accountability systems argue that there are costs associated with not having an assessment system. Researchers have described a number of benefits of accountability and testing (Bishop, 1997; Cizek, 2001). In the aggregate, test results can inform school officials and teachers about the value of different curricula and instructional materials so that resources can be better directed. At the individual student level, they can be used for screening, diagnosis and progress monitoring and can help guide instruction and prevent academic failure, particularly for low-performing students (Herrington, 1996). Moreover, a state-mandated test is likely to be the most valid, reliable test a student takes all year (Cizek, 2001).

In the absence of standardized assessments, public officials, the general public and parents lack any yardstick by which to measure how well their schools are educating their students. Affluent parents have more time, information and other resources by which to monitor their children and their schools' performance. That is to say, children from more affluent communities already have an informal accountability system in place—informed, focused parents. In the absence of standardized testing, poor children

are likely to suffer the most. An assessment system that forces attention on these children and their schools—and whose costs are not exorbitant—might well be a good investment. Secondly, greater transparency offered by public reporting of standardized test scores may make the public less concerned about poor management of public schools and more comfortable with higher levels of expenditures for K-12 education.

Critics of accountability, on the other hand, paint another scenario. Pressure to produce students that test well on standardized assessments and to use prescribed curricula designed uniquely for that purpose may make the teaching profession less attractive, accelerating already dangerously high numbers of teachers choosing to leave the profession. It may also exacerbate the existing state of affairs where the best teachers are found in the schools with the strongest students. Increased assessments tied to greater accountability may result in even more sorting out of teachers (Hanushek et al., 2001). That is to say, the more effective teachers may increase their pressure on the system to allow them to transfer to schools whose students score better. The danger is that the impact of the assessments may be to worsen the fundamental problem that accountability is trying to solve: raising the achievement of the lowest performing students. Increasing transfers out of the system—or increasing within-system transfers of the better teachers to the least needy children—could raise the costs associated with retention and remediation.

Conclusions

Many states are facing financial strain caused in part by the general public's growing demand for public services, including education, and a reluctance to raise taxes. In this climate of resource scarcity, all new policies and programs face enhanced scrutiny and evaluation. How much do they cost? What are the benefits gained? Florida, like many other states, is grappling with these issues around its assessment system, the FCAT, and its accountability system, the *A+ Plan*. A tenet of sound management of any enterprise is evaluation of effectiveness. Some expenditure of funds is required to perform that function. The question is what level of funds, and for what kind of evaluation?

As the analysis in this brief has shown, estimating the cost of accountability and assessment in Florida and elsewhere is complex. The direct cost of accountability could be minimal compared to overall costs, as several analyses have demonstrated. But any assessment of direct costs is likely to be incomplete. Indirect costs or the cost of opportunities not pursued could be very high. It is certainly important not to view costs alone, but rather in relation to benefits yielded. In conducting a sound estimation, the analysis must take into account different assumptions and factors affecting direct and indirect costs and provide a range of estimates covering most of the known variations.

References

- Bishop, J. H. (1997). The effect of national standards and curriculum-based exams on achievement. *American Economic Review*, 87(2), 260-264.
- Cizek, G. J. (2001). More unintended consequences of high-stakes testing. *Educational Measurement: Issues and Practice*, 20(40), 19-30.
- Figlio, D. (2002). *Funding and accountability: Some conceptual and technical issues in state aid reform*. Paper presented at Syracuse University's Conference on State Aid to Education, Syracuse, NY.
- Figlio, D., & Getzler, L. (2002). *Accountability, ability and disability: Gaming the system* (NBER Working Paper No. 9307). Cambridge, MA: National Bureau of Economic Research.
- Florida Department of Education (2004, May). *Assessment and accountability briefing book*. Retrieved November 28, 2004, from <http://www.firn.edu/doe/sas/fcat/pdf/fcataabb.pdf>
- Haney, W., Madaus, G., & Lyons, R. (1993). *The fractured marketplace for standardized testing*. Boston, MA: Kluwer Academic Publishers.
- Hanushek, E., Kain, J., & Rivkin, S. (2001). *Why public schools lose teachers* (NBER Working Paper No. 8599). Cambridge, MA: National Bureau of Economic Research.
- Herrington, Carolyn D. (1996). The politics of school-level finance data and state policy making. In L. O. Picus and J. L. Wattenbarger (Eds.), *Where does the money go? Resource allocation in elementary and secondary school* (pp. 236-252). Thousand Oaks, CA: Sage.
- Hoxby, C. (2002). *The cost of accountability* (NBER Working Paper No. 8855). Cambridge, MA: National Bureau of Economic Research.
- Monk, D. (1994). *The costs of systemic education reform: A summary report*. Ithaca, NY: Cornell University, Department of Education.
- Phelps, R. (2000). Estimating the cost of standardized student testing in the United States. *Journal of Education Finance*, 24, 343-380.
- Picus, L. (1994). *A conceptual framework for analyzing the costs of alternative assessment* (CSE Technical Report 384). Los Angeles: CRESST/University of California-Los Angeles.

United States General Accounting Office (2003). *Title-I: Characteristics of tests will influence expenses; Information sharing may help states realize efficiencies* (GAO-03-389). Washington, DC: GAO.