DOCUMENT RESUME

ED 376 210 TM 022 328

AUTHOR	Monk, David H.
TITLE	The Costs of Systemic Education Reform: Conceptual Issues and Preliminary Estimates.
SPONS AGENCY	Consortium for Policy Research in Education, New Brunswick, NJ.; John D. and Catherine T. MacArthur Foundation, Chicago, IL.; Pew Charitable Trusts, Philadelphia, PA.
PUB DATE	Nov 93
CONTRACT	R1178G10039
NOTE	240p.; Paper presented at the Annual Meeting of the American Educational Research Association (New Orleans, LA, April 4-8, 1994).
PUB TYPE	Reports - Evaluative/Feasibility (142) -- Speeches/Conference Papers (150)
EDRS PRICE	MF01/PC10 Plus Postage.
DESCRIPTORS	Cost Effectiveness; *Cost Estimates; *Educational Change; Elementary Secondary Education; *Estimation (Mathematics); Policy Formation; *Program Development; Program Implementation; Research Design; State Programs; Statistical Analysis; *Systems Development
IDENTIFIERS	Authentic Assessment; Large Scale Assessment; *New Standards Project (LRDC); Performance Based Evaluation; Reform Efforts

ABSTRACT

An overview of cost analysis is provided as it relates to a particular educational reform--the use of performance or authentic assessment on a large scale as a means of transforming entire educational systems. The focus of the inquiry is the New Standards Project (NSP), a joint effort of the National Center on Education and the Economy and the Learning Research and Development Center at the University of Pittsburgh. The design of the NSP is the prototype of the cost analyses discussed, but the costs of the NSP are not discussed per se. Chapter 2 provides a conceptual examination of pitfalls associated with cost analysis. Chapters 3 through 6 generate preliminary estimates of the costs associated with large-scale pupil-performance assessment. While the Chapter 3 focus is on development costs, Chapters 4, 5, and 6 deal with operations costs in large, medium, and small states, respectively. Chapter 7 draws the results together and places cost estimates in context, placing upper and lower bounds on the magnitude of costs associated with large-scale performance assessment. Systemic reform does not lend itself to conventional cost analysis, but these results are intended to provide some guidelines for policymakers. Two figures and 11 tables illustrate the analysis. (Contains 37 references.) (SLD)

THE COSTS OF SYSTEMIC EDUCATION REFORM: CONCEPTUAL ISSUES AND PRELIMINARY ESTIMATES

David H. Monk
Department of Education
Cornell University

Support for this research was provided by the McArthur Foundation and the Pew Charitable Trusts through grants supplied to the New Standards Project. Support was also provided by the Finance Center of the Consortium for Policy Research in Education (CPRE), a consortium of the University of Southern California, Rutgers University, Cornell University, Harvard University, Michigan State University, Stanford University and the University of Wisonsin-Madison. The CPRE support was derived from grant #R1178G10039 from the U.S. Department of Education, Office of Educational Research and Improvement. The views expressed are those of the author and are not necessarily shared by the the New Standards Project or its sponsors, CPRE or its partners, or the U.S. Department of Education.

November 1993

CHAPTER 1

INTRODUCTION AND ACKNOWLEDGEMENTS

Cost analyses, particularly as they apply to evaluation in education, are of relatively recent origin and are not widespread (Catterall, 1988; Haller 1974; Levin 1991; Monk and King, in press). Various reasons have been offered for the apparent neglect, including the absence of appropriate training (Levin, 1991) as well as deeply rooted conceptual and data problems that interfere with analysts' ability to draw the straightforward conclusions sought by policymakers (Monk and King, in press; Thomas 1990). There is, nevertheless, no denying the salience of policymakers' interest in costs, and some impressive methodological progress has been made (see, for examples, Barnett 1985, 1991; Jamison, Klees, and Wells 1978; and Levin, Glass, and Meister 1984).

In this study, I provide an overview of cost analysis as it pertains to a particular educational reform--the advent of performance or authentic assessment on a large scale as a means of transforming entire educational systems. I use as the focus of my inquiry the New Standards Project (hereafter, NSP), a joint effort of the National Center on Education and the Economy and the Learning Research and Development Center at the University of Pittsburgh (NSP 1992). By organizing the discussion around a particular instance of reform, I seek to make the analysis

relatively concrete and useful to policymakers faced with decisions about whether and how to proceed with pupil performance assessment as a major component of school reform initiatives.

While it is true that the design of the New Standards Project is the prototype for the cost analyses I conduct, it has not been my goal to assess the costs of the NSP, per se. My goal is broader, since I seek to throw light on the cost implications of large-scale pupil performance assessment as a vehicle for achieving what is becoming known as systemic school reform.[1] In part, this decision to broaden the analysis is pragmatic, since the design of NSP itself is evolving and any attempt to "cost-out" its components risks being significantly out-of-date as soon as the analysis is complete. In part, the decision is to provide additional insight into the costs of pupil performance, since it is possible that practices that evolve will diverge significantly from the NSP model. I make explicit note of the departures I make

---

[1] A large literature has developed around this approach to reform. The approach has several components including: (1) curriculum frameworks that specify what students need to learn; (2) coherent state and local policies designed to enhance the teaching and learning of what is spelled-out in the curriculum frameworks; (3) new governance systems that achieve accountability by fostering flexibility and control at the school site coupled with refined pupil assessment mechanisms that provide relevant feedback that can be used for a variety of purposes at a variety of levels within educational systems. Much debate surrounds the use of these pupil assessment mechanisms. It is nevertheless clear that pupil assessment, however it is used, is central to the systemic reform movement within education, and for this reason warrants careful scrutiny by policymakers at federal, state, as well as local levels of school governance. For more on systemic reform see the collection of papers edited by Furhman (1993), especially the papers by O'Day and Smith (1993) and Clune (1993).

from the NSP model.

The study has two major components. Chapter 2 provides a conceptual examination of pitfalls associated with cost analysis. In particular, I anticipate problems a cost analyst is likely to encounter when faced with the task of estimating costs of pupil performance assessment, and offer suggestions about how to respond. Chapters 3-6 comprise the second major component of the study where I generate preliminary estimates of the costs associated with large-scale pupil performance assessment. The focus in Chapter 3 is on development costs; Chapters 4, 5, and 6 deal with operations costs.

Because education is a state responsibility, the operations cost estimates I generate are state specific. Each of the three chapters devoted to operations costs is tailored to a different sized state: large, medium and small, respectively. Each of these chapters begins with a description of the relevant state and proceeds to derive the associated costs. The chapters parallel one another closely, and most readers will find it sufficient to concentrate on the chapter dealing with the type of state in which there is the greatest interest.

The study concludes with Chapter 7 where I draw together the results and place the cost estimates in context. My primary goal is to provide policymakers from a variety of states useful information that will inform decisions that must be made in the near term about the viability of large scale performance assessment as a major vehicle of education reform.

These cost estimates depend heavily on a series of assumptions, and this is clearly problematic. The standard defense applies: I am as explicit as possible about the assumptions I make, and I invite the reader to alter them. I have also explored the consequences of making a range of assumptions, some more favorable to the proponents of performance assessment than others. As I have indicated, my cost estimates are intended to place upper and lower bounds on the magnitude of costs associated with large scale efforts to introduce performance assessment into K-12 education within the United States.

The study has benefitted from the assistance offered by many individuals and organizations. Funding was provided by The Pew Charitable Trusts and the John D. and Catherine T. MacArthur Foundation. Additional support was provided by the U.S. Department of Education through the Educational Finance and Productivity Center that is operated by the Consortium for Policy Research in Education (CPRE). The Learning Research and Development Center at the University of Pittsburgh and the National Center on Education and the Economy jointly administer the NSP and provided able assistance on numerous occasions. The individuals who have provided counsel include Susan Bennett, Dominic Brewer, James Fox, James Gilchrist, Emil Haller, Daphne Hardcastle, Jennifer King, Archie Lapointe, William Lepley, Allan Odden, Lawrence Picus, Dan Resnick, Lauren Resnick, Christopher Roellke, and Marc Tucker. I am very grateful for the help and encouragement offered by these individuals. The views expressed

are my own and whatever errors remain are, of course, my responsibility.

CHAPTER 2

CONCEPTUAL ISSUES

The seemingly straightforward interest in estimating costs gives rise to a large number of significant conceptual problems. This chapter provides an overview of these problems and begins with a discussion of the important distinction that needs to be made and maintained between expenditures and costs. Much confusion stems from a lack of clarity here, and it is therefore a useful point of departure. Next comes an examination of issues that arise once an analyst has begun a bonafide cost analysis. These include the identification of relevant foregone opportunities and their measurement; the handling of ambiguous costs; the allowance for the fact that costs can be very unevenly imposed across categories of actors within the system under study; the selection of the appropriate unit of analysis; and the appropriate adjustment for economic phenomena such as diminishing marginal rates of productivity. These points are drawn upon in Chapters 3-6 where attention turns to the trial cost analyses for large-scale pupil performance assessment reforms.

Distinguishing Between Costs and Expenditures

Costs are measures of what must be foregone to realize some benefit, and for this reason they cannot be divorced from benefits. Expenditures, in contrast, are measures of resource

flows regardless of their consequence. A cost analysis requires a comparison of benefits; an expenditure analysis does not. The cost of pursuing one activity rather than another is the highest benefit foregone of devoting resources to the activity in question.[1]

Information about expenditures is generally more readily available than information about costs.[2] We hire armies of accountants to keep track of expenditures; there is no comparable corps of cost analysts. This is particularly true in education where knowledge of costs is impeded by the multiplicity of possible benefits coupled with a rudimentary knowledge of how resources are translated into educational outcomes (Monk 1992). In short, there is no viable means of distinguishing between expenditures that are required given present technology and those that are due to inefficiency and waste.

The difficulties are only compounded when the goal is to estimate costs in an unexplored aspect of education such as the performance assessment of students. Ignorance about the

---

[1] An extensive literature has grown around the conceptualization of costs. For examples of quite thorough treatments see, Bowman (1966); Buchanan (1969); Thomas (1990). For a more accessible introduction, see Walsh (1970). For a good and nontechnical overview of cost analysis as it applies to evaluation, see Haller (1974).

[2] While this is true in a relative sense, it is remarkable to observe how limited our actual ability is to keep track of expenditures for education. See Fowler (1992) for a discussion of the gaps in the federal government's school finance data collection.

production realities surrounding performance assessment is widespread if for no other reason than the fact that many of the initiatives are still being designed or are at very early stages of implementation (Pelavin, 1992). Moreover, the number of goals being pursued by performance assessment reforms is remarkably large. A review of the New Standards Proposal (1992) reveals no fewer than nine such goals, some of which have the potential to be contradictory.[3] A serious commitment to estimating the costs of performance assessment must involve determining the resources necessary to accomplish these numerous goals and their best alternative use. Anything short of this is an exercise in estimating expenditures.

Unfortunately, the more readily available expenditure data are of limited use for policymaking. They can be useful if a

---

[3] Here is a list of the various things that the New Standard Project is seeking to accomplish (New Standards Proposal, 1992): fundamentally change what is taught and learned; raise expectations that teachers have of students; greatly increase student motivations and effort; raise student performance across the board; substantially close the gap between the best and worst performers; reward student effort to master a thinking curriculum by providing access to college and jobs to those who do so; reward school professionals who helped their students succeed against the new standard; inform parents and the public of the standards to which students would be held and the material they were expected to master; and establish national standards but retain local initiative and creativity. If the desire to raise student performance across the board translates into a desire to raise the mean level of achievement, there can arise a contradiction with the simultaneous desire to close gaps between the best and worst performers, assuming the resource base is finite.

decision has been made to proceed with a project and the question is whether there are sufficient resources identified for implementation, or if there is curiosity about how much was spent on a particular activity. But expenditure data are quite useless if the more fundamental question is being asked about whether or not or how to proceed with a project. What makes matters worse is that expenditure data can masquerade as cost data and be misused in policymaking.

For example, if an analyst were to provide expenditure estimates associated with two approaches to pupil assessment, compare them head-to-head, and use the results to draw conclusions about how much more the one approach "costs" relative to the other, the analyst would be assuming implicitly that the two assessments are intended to accomplish the same goals and are each afflicted to the same degree with inefficiency. Only under these conditions would the comparisons be valid and have relevance for a decision about whether to do more or less of one or the other type of assessment. In cases where these demanding conditions do not hold, the comparisons are not valid and can be seriously misleading.

This point can be further illustrated by examining a recent instance where expenditure data were cited in a cost context for the purpose of questioning the viability of relying more heavily on performance assessment for students in U.S. schools. Theodore Sizer, in a forum sponsored by Education Week, suggested that George Madaus' research indicated that the dollar costs of "truly

authentic assessments" range between 6 and 20 times as much as current practice (Education Week June 17, 1992, pg. S4). Sizer used these figures to caution reformers about the potentially high costs of authentic assessment. He went on to make the quite sensible point that costs need to be taken seriously since they represent a host of alternative reforms that might otherwise be pursued. I have no quarrel with Sizer's larger point about the importance of looking at costs. However, it would appear that the figures he cites are based on expenditure data and that he is overstating what we know about costs.

A closer look at what Madaus said about the costs of assessment is instructive. His observations occur in the context of a study he and a colleague, Thomas Kellaghan, conducted of student examinations systems in Europe. Among their findings is information about what Ireland and the United Kingdom spend on their external examination system (Madaus and Kellaghan, 1991). Specifically, they report a figure of $107 per examined student for Britain and Ireland, and estimate that if Massachusetts were to adopt one of these models to test its comparably aged students (16 year olds), the cost would be almost $7 million. These authors then compared this figure with the $1.2 million they claim Massachusetts currently spends to test the reading, writing, and arithmetic achievements of students at three grade levels (using machine scoring for the reading and mathematics tests), and concluded that were Massachusetts to adopt a European model of

external exams, there would be "very substantial financial implications" (Madaus and Kellaghan, 1991, pg. 22).

What Madaus and Kellaghan report are differences in expenditures across quite different types of assessment efforts. They are correct to conclude that expenditures in Massachusetts would rise if the European model were adopted, but their figures cannot be used to conclude that the European model costs more, or that authentic assessment costs more than traditional assessment. The two approaches to assessment are fundamentally different and the respective expenditure levels are not strictly comparable.[4]

Discerning Costs

Having distinguished between expenditures and costs, we can take the next step and examine issues that need to be resolved before a cost analysis of performance assessment can proceed.

A. Identifying Relevant Foregone Benefits

---

[4] There have been a number of other attempts to make estimates of resource outlays for one or another type of assessment program. For example, Bauer (1992) surveyed Test Directors and estimated the average annual costs of testing per pupil to be $4.79. Haney, Madaus, and Lyons (1993) estimated a direct outlay of less than $.80 per student per test hour. The Office of Technology Assessment compiled a state-by-state listing of the costs of State Assessment Programs and reported that costs in 1988 dollars ranged from $1.12 to $39.42 per student (as cited in Haney, Madaus, & Lyons 1993, p.111). Finally, the General Accounting Office recently estimated that systemwide testing costs about $15 per student (USGAO 1993).

Costs cannot be defined in the absence of alternatives. Costs are incurred to the degree that some desirable alternative is foregone and the associated benefits are not realized. Thus, when a resource is devoted to one use, the benefits associated with all of the alternative possible uses of the resource are relevant to the task of determining the resource's cost.

Possible restrictions on the range of alternative uses. Which among all the possible uses is the relevant alternative use? Textbook definitions of opportunity costs identify the relevant alternative use as the best alternative use, but this is not always helpful since considerable ambiguity can surround what counts as "best."[5] An example can make this point clear.

Suppose the task is to determine the cost of time a student might spend attending a Friday evening basketball game. By definition, the opportunity cost of the student's time is the "best" opportunity foregone by virtue of spending the Friday evening at the basketball game. The pertinent question concerns the broadness of the relevant range of alternative opportunities. Suppose the student in question is under close parental supervision so that the only alternative to going to the basketball game is spending a quiet evening at home, and let us

---

[5] The Office of Technology Assessment (U.S. OTA 1992, p. 27) speaks more generally about the "value of foregone alternative action," and risks generating confusion. It is not just any foregone alternative action that corresponds to the cost. It is, instead, the best or more highly valued alternative action.

suppose further that this is not a very attractive alternative use to the student.[6] Under these conditions, the cost of the time spent at the basketball game (from the student's perspective) is quite low--not much is being foregone.

Now suppose that the conditions are different and the range of alternative choices is broadened to include going to a jolly party with really keen people. Assuming this is an attractive alternative use (again, from the perspective of the student), the cost of attending the basketball game has gone up, perhaps dramatically. We have reached two quite different conclusions about a cost, depending on how broadly we choose to define the relevant range of alternative uses.

This variability in the range of relevant alternatives can have bearing on our interest in establishing cost estimates for performance assessment. If we ask the question: "What is the cost of resources that are devoted to performance assessment activities?," the textbook answer will be: "The benefits of the best possible alternative uses to which these resources might have been put." This answer links the cost of performance assessment to the benefits of any conceivable alternative reform (within as well as outside of education). The more beneficial the alternative use(s), the more costly it becomes to devote resources to performance assessment.

However, there may also be a sense in which the range of

---

6 Indeed, the parents' supervision could be so close that the student is not even aware of a host of alternative uses.

alternative uses to which the resources required for performance assessment might be put is more severely constrained. Suppose, for example, that the only relevant alternative use for resources being devoted to performance assessment is conventional assessment. If this is the case, the costs of performance assessment will be measured in terms of the benefits of conventional assessment that are foregone. And to the degree that the benefits of conventional assessment are more modest than those associated with other possible uses, the costs of performance assessment will be lowered by virtue of the restriction on the range of relevant alternatives.

Why would it be appropriate to restrict the range of alternative uses? One justification could be based on behavioral expectations. If it is likely that performance assessment will substitute for conventional assessment, then there is a sense in which the costs of devoting resources to performance assessment come at the expense of fewer resources going toward conventional assessment. Some data are beginning to appear that examine the degree to which new assessment approaches substitute for existing assessment efforts. For example, the U.S. General Accounting Office (1993, pg. 44) reports that 41% of the districts they surveyed substituted a state provided test for local tests despite the fact that in the district's opinion the tests were quite dissimilar. In cases where the district thought the tests were similar, over 80% reported making the substitution.

However, assuming conventional assessment is not the best possible alternative use of the performance assessment dollars, it follows that foregone conventional assessment benefits are understating the true economic costs of performance assessment. The point is that a decision needs to be made about what counts as the relevant foregone use.

Sources of variation in benefit levels. It is important to be more specific about the dimensions along which foregone benefits can vary. They derive from two sources.

First, there is the direct contribution to the relevant decision maker's sense of well-being. It is a question of how well aligned the alternatives being foregone are to the relevant decision maker's preferences. Of course, this presumes clarity about who the relevant decision maker is. Suffice it to say that views about how valuable different foregone benefits are can vary substantially among those playing different roles.[7]

The basketball example can help to illustrate this dimension of the valuation problem. Going to a party with a given set of characteristics contributes in a particular way to the student's sense of well-being. This may be high or low or in-between, and it depends on how the student feels about parties. The more important party-going is to the student, the more costly it

---

[7] For more about how it is reasonable for different actors within educational systems to disagree fundamentally over the value of a central resource such as student time, see Monk (1982).

becomes for the student to spend the time at the basketball game, assuming s/he is aware of the party option.

Second, there is also a productivity dimension to consider. Parties can be good or bad, jolly or not, and our student's sense of the cost of going to the basketball game will be affected by his/her perception of the level at which the party will operate.[8] In other words, it may be the case that a party has the potential to be very beneficial in the student's mind, but the reality could be quite different.

Again, there is a parallel with the problem of assigning costs to performance assessment. The foregone alternative used to assign value to the performance assessment resources may or may not be contributing benefits that are highly valued by the society. In other words, the benefits being produced may not align very well with what the society is seeking.[9] Moreover, the

---

[8] This concern over the level of production is conceptually distinct from a concern over how efficiently the party is produced. The student is less likely to be concerned about how efficiently resources are being transformed into party outcomes, largely because the resources are presumably coming from others. Even if we recognize that a party-going student will eventually be expected to host a party and thereby incur costs, it is not obvious that the student will be concerned about efficiency per se. Just because the student's associates run inefficient parties (and expend more resources than are necessary), does not mean that the student needs to follow suit.

[9] If the relevant alternative is conventional assessment, it could be the case that conventional assessment places too much emphasis on rote learning and lower cognitive capabilities. It could be the case that conventional assessment (assuming this portrayal is accurate) is ill-serving the interests of society as we move into the 21st century.

alternative use may or may not be operating at a very high level. Serious inefficiencies may be limiting production of the relevant benefits.

It follows that misalignment between the alternative use and the society's priorities as well as inefficiency in the production of the relevant alternative benefits have implications for the cost of performance assessment. This makes sense intuitively. It costs less to replace a poor practice than it does to replace a good practice. However, this kind of thinking begs the question about whether the poor practice could be improved. It also sidesteps the possibility that the restriction on the range of relevant alternatives is artificially drawn.

Lumpiness. Costs can be conceived of at the margin (i.e., the cost of devoting additional resources to a given use) or in a cumulative sense (i.e., the sum of benefits foregone given the allocation of some bundle of resources in a given direction). One reason why the two types of costs may differ stems from the potential for the alternative uses to be lumpy in their nature. In the basketball game example, the game may take more time than the alternative party. Thus, the cost of the time devoted to the game needs to be valued in terms of the benefits of the party plus the benefits of the best alternative use of time following the party. And in the case of performance assessment, the resources devoted to performance assessment may be greater than those devoted to the relevant alternative use, say conventional assessment. Under these circumstances, the cost of performance

assessment consists of the foregone benefits of conventional assessment plus whatever benefits are foregone because of the additional resources devoted to performance assessment.

B. Implications for Measurement

Measurement questions quickly crowd discussions about foregone benefits or opportunities. Recall that the textbook definition of an opportunity cost makes reference to the best benefit foregone, not the most easily measured benefit foregone. And yet, cost analysts are under considerable pressure to develop metrics for the benefits they are assessing. A common strategy is to rely on market valuations of foregone benefits despite the fact that these dollar measures may not reflect the most highly valued foregone benefits.

The Friday night basketball game example can also help clarify this issue. Both alternative uses of the student's time that we considered above (spending the time at home or at the party) do not lend themselves to a dollar metric. There is, however, a third alternative use that is relatively easy to cost in dollars--namely, the wage the student could command if the student spent the evening working. While this alternative use may be relatively easy to measure, it could be a very misleading cost estimate for the simple reason that it is hardly obvious that it represents the "best" alternative use in the student's mind.

The distinction between easy and hard to measure benefits has

relevance for assigning costs to performance assessment. It would be desirable to have direct measures of the net benefits associated with the best alternative being foregone because of the proposed shift toward performance assessment. However, such measures are not readily available and would require a major effort with no guarantee of success. A second-best strategy involves accepting the claim that the net benefit of the alternative use can be measured by the dollar value of the resources devoted to it. If this strategy is pursued, an important part of analyzing the costs of performance assessment becomes the calculation of expenditures on the best alternative use(s) to which the resources might be put. But, this is equivalent to calculating the dollar value of the resources devoted to the intended use, and the result is the use of either actual or anticipated expenditures on the intended use as the measure of the relevant costs. This approach to estimating costs is sometimes called the "ingredients" approach or method. It places a heavy emphasis on using expenditures to measure costs and can thereby contribute to the confusion surrounding the very important conceptual difference between the two.[10]

The use of expenditures to measure costs has some merit. After all, dollars are broadly instrumental and their expenditure on a given ingredient does provide a measure of all the market based opportunities that are being sacrificed by virtue of the

---

[10] For a good overview of the "ingredients" method and its application to program evaluation, see Levin (1983, pp. 51-59).

decision to spend. But, the underlying prices which give meaning to the dollar measures, are generated by markets, and markets can vary widely in how well they function. Where markets do not function well, it is possible for the dollars spent on ingredients to be quite unrelated to actual benefits derived.

From a neoclassical economist's perspective, markets do not function well when they operate in non-competitive environments. In the case of education, the deep involvement of the state is viewed by some as a serious limit on how well education markets can succeed at efficiently producing the correct mix of educational outcomes.[11] If these critics are correct and if resources devoted to performance assessment will come at the expense of resources devoted to other educational uses, then the use of the ingredients method to estimating the costs of performance assessment risks overstating the relevant costs. In other words, under these assumptions, totaling the dollars that will need to be spent on performance assessment would overstate the opportunities society would truly forego if performance assessment were implemented.

The point is not to debate the merits of public intervention in the functioning of education markets. Rather, the point is to recognize that the use of the ingredients method will overstate the costs of performance assessment to the degree that misalignment with social priorities and inefficiency in production

---

[11] See, for example, Chubb and Moe (1990).

characterize the relevant alternative use of resources that could otherwise be devoted to performance assessment.

Figure 1 illustrates this point. It shows three possible conceptualizations of the costs of performance assessment. In panel A, the assumption is that it is reasonable to assign costs to performance assessment that correspond to the anticipated expenditures associated with performance assessment, and the costs of performance assessment are represented by OC.

Figure 1

Alternative Conceptualizations of Cost

```
          A                    B                    C
0     |         |         |         |         |         |
      |         |         |         |         |         |
      |         |         |         |         |         |
      |         |         |         |         |         |
      |         |         |         |         |         |
C**   |         |         |         |         |_________
      |         |         |         |
C*    |         |         |_________
      |         |
      |         |
C     |_________
```

Panel B, reflects a presumption that the dollar value of the

expenditures overstates the true costs. The reason for the overstating stems from a presumed lack of congruence between what the alternative contributes to social welfare and what is truly desired. To be more concrete, if the relevant alternative is conventional assessment, the presumption is that conventional assessment is running efficiently but is producing a less than optimal mix of outcomes. In other words, the dollars devoted to conventional assessment could generate a mix of more highly valued benefits. The associated costs are OC* and OC* < OC.

In panel C, the lack of congruence idea is carried forward and a degree of production inefficiency is added. The idea here is that not only are the foregone benefits not very well aligned with social preferences, they are not being produced at a level that is technically possible. This further reduces the cost of the rival program, since less is being lost if the change were to be made.[12] For Panel C, the cost of performance assessment is OC** and OC** < OC* < OC.

These arguments pertain to questions about the costs associated with performance assessment. If we alter the question slightly and ask how much more it would cost to implement a system of performance assessment within an existing school system, there

---

[12] Whatever inefficiencies exist within performance assessment will be introduced by virtue of the inclusion of unnecessary ingredients. Whatever misalignments might exist between what performance assessment contributes and what society is seeking will not be captured by this kind of cost analysis. Instead, a benefit-cost analysis would be required and the misalignment would enter on the benefit side of the analysis.

is an additional phenomenon to consider--namely, the possible absorption of performance assessment costs.

Costs will be absorbed to the degree that the performance assessment reform substitutes in practice for some aspect of the status quo. For example, to the degree that performance assessment can substitute for conventional assessment and existing staff development efforts, the marginal cost of implementing performance assessment will be diminished.

There is, however, an important difference between the degree to which one use of resources can substitute for another and the likelihood that the substitution will actually take place in practice. The complex decision making processes that give rise to actual practice in schools are difficult to assess and involve important political as well as economic phenomena. This mixing of political and economic phenomena gives rise to some ambiguity about the relevant costs. From a strict economic perspective, the cost is the best alternative foregone, regardless of what happens in practice. But, from a policymaking perspective, the potential for substitutions to take place is clearly relevant and has bearing on both the estimates of costs and their subsequent use in policy debates.

An important question that is much easier to ask than to answer concerns the degree to which misalignment with social goals and/or inefficient production of one resource use enhances the likelihood of substitution with an alternative. In the present context, the question is about the degree to which misalignment

and inefficiency associated with conventional assessment is likely to enhance the prospects of substitution in practice with performance assessment. If this kind of link exists, it follows that misalignments and production inefficiencies have bearing on two aspects of cost: (1) the cost of the resources required for the reform; and (2) the cost of implementation. Figure 2 illustrates both of these cost components.

Figure 2

Alternative Conceptualizations of Adding

Performance Assessment to an Existing Educational System

```
          A                    B                    C

      |         |          |         |          |         |
      |         |          |         |          |         |
      | REG     |          | REG     |          | REG     |
      | INST    |          | INST    |          | INST    |
      |         |          |         |          |         |
      |         |          |         |          |_________|
      |_________          |_________          |         |
      |         |          |         |          CONVEN
      | CONVEN  |          | CONVEN  |          |__ASSESS__
      | ASSESS  |          | ASSESS  |          | PERF    |
0     |_________          |_________          | ASSESS  |
C***                       |         |          |_________
                           | PERF    |
                           | ASSESS  |
C                          |_________
```

Panel A in Figure 2 represents a schooling system before the advent of performance assessment. The figure includes an admittedly artificial distinction between the costs of regular instruction and the costs of conventional assessment. Panel B reflects the addition of the performance assessment reform where the costs are valued in terms of the full dollar value of the resources required for performance assessment and where performance assessment is considered a complete add-on to existing practices. The magnitude of this cost, 0C, in Panel B is the same as that depicted in Panel A of Figure 1. In Panel C, two things have happened: (1) there has been an adjustment to reflect the presumption that the dollar value of the resources required for performance assessment overstate the cost (this is the same adjustment made in panel C of Figure 1) and (2) an allowance has been made for the absorption of some portion of the costs of performance assessment into the costs of both the regular instructional program and the conventional assessment program. In other words, a substitution is presumed to have taken place between what was in place and the performance assessment reform. The figure is drawn to suggest that these two adjustments have a significant impact on the costs associated with performance assessment.

These arguments suggest that the conventional ingredients method can overstate the true economic costs of a reform like pupil performance assessment, but they offer little guidance about

the magnitude of the overstatement. A case can be made for making an offsetting adjustment, but for these offsets to be credible, there needs to be reason to believe that the proposed new use (performance assessment in this case) will be less likely to suffer from both a misalignment with social welfare interests and inefficiency in production.

It is probably easier to make the better alignment case for performance assessment than the productivity case. There appears to be consensus that the kinds of human performance dealt with by performance assessment are likely to become more and not less important to economic as well as social functioning as time passes (Marshall and Tucker 1992). However, it is hardly obvious that so-called conventional assessment has no role to play in assessing these kinds of capabilities.

The productivity case is even more difficult to make since the reform scenario envisioned within the NSP keeps the public school governance system largely intact. If the existing governance system gave rise to inefficiency within the conventional assessment program, what reason is there to expect performance assessment to suffer a different fate? Perhaps the sometimes parallel efforts to restructure school governance and to more directly involve teachers and parents will have salutary effects, but this is speculative at best.[13]

I cannot resolve these matters here, and I choose to respond

---

[13] See O'Day and Smith (1993) for more on the kinds of governance changes that are part of systemic reform initiatives.

to the problem by calculating costs according to different assumptions about the magnitudes of the relevant offsets. In particular, I make several explicit assumptions about the magnitudes of the offsets and include the case where the offset is zero. Indeed, the zero offset case where costs are estimated on the basis of projected expenditures on ingredients will be the starting point for the analysis.

C. Handling Ambiguous Costs

Ambiguous costs involve real but in some sense unnecessary expenditures of resources. In a strict sense, they are not costs, since they are not necessary to accomplish some end. In another sense, they are quite real to the extent that those involved perceive the expenditures to be necessary.

The importance of these costs arose in conjunction with a cost analysis of the Texas Examination of Current Administrators and Teachers (TECAT). Shepard and Kreitzer (1987) drew attention to the issue when they showed that their cost estimates of the TECAT went up dramatically when they included a valuation of the time teachers devoted to preparing for the test. It is at least arguable that such preparation time was not intended by the state to be necessary. Nevertheless, teachers spent the time, and the time required them to forego opportunities. Resources were expended, and the question is whether or not to treat them as costs. It is possible for the new performance assessments to

generate significant costs of this kind, particularly if the stakes associated with the test are high.

In the empirical analyses to follow, I make varying allowances for the presence of these costs through the use of alternative cost scenarios. The "best" case scenario provides the smallest allowance for ambiguous costs; the "worst" case scenario reflects the assumption that these costs are substantively important.

D. Defining the Locus of Costs

It is also important to be clear about whose perspective is being considered in a cost analysis, since the imposition of costs can vary widely across categories of actors within educational systems. An analyst might show that costs of a reform are relatively modest at the state level (or from a funding agency's perspective). Armed with these results, policymakers might go ahead and implement the reform only to discover subsequently that the neglected costs borne by actors located at other levels of the system were sufficiently large to thwart the entire reform.

Shepard and Kreitzer (1987), for example, found that the contracted resource commitment for the teacher examination at the state level was on the order of $5 million dollars, but estimated that the total tax support for the program amounted to more than $35 million when local costs were included. The Office of Technology Assessment (1992), hereafter OTA, also found a large

discrepancy between the estimated outlays for a conventional standardized testing program (including: contracted materials and services as well as district testing personnel) and a more comprehensive estimate of the outlays which took account of the time teachers spend preparing students for and administering the examination. The OTA estimates ranged between $6 per student per test administration and $110 per student per test administration, and illustrate how sensitive the results can be to decisions about what to include and exclude.

As further evidence of the importance of being attentive to the locus of costs, consider OTA's analysis of school districts' likely behavioral responses to alternative types of assessment programs. OTA distinguished between one hypothetical testing program that costs little in terms of direct dollar outlays but is quite costly in terms of the costs imposed on students, what OTA calls opportunity costs. By assumption this testing program (Type I) has little or no instructional value. Whatever time a teacher spends preparing students for this type of test requires a like amount of time to be withdrawn from productive instructional uses. The alternative (Type II) program has the opposite features: it is costly in terms of direct costs but has minimal opportunity costs. This corresponds to a program where the development of assessment tasks and their subsequent scoring are quite costly but where the assessment fits very nicely with instruction and even complements teachers' efforts to teach. Whatever time a teacher devotes to preparing students for this type of test has no adverse effect on

learning.

According to OTA, the costs of the Type I test start low and increase as more time is devoted to assessment, while the costs of the second option are constant and do not vary with the amount of time devoted to the assessment. OTA identified a cross-over point where the initially lower costs of Type I meet and then go beyond the costs of Type II, and claimed that at the cross-over point the district (emphasis added) would be indifferent between the two testing programs.

This conclusion misses an important point about who bears what cost. To the degree that students bear the opportunity costs associated with the Type I assessment program, why would the district care about these costs? My conclusion is different from OTA's: In my view, at the crossover point, the district would still prefer to use the Type I assessments. The opportunity costs, which are assumed to be large and real, are imposed on students who are limited in their ability to organize and make their needs known. In sharp contrast, the additional direct expenditures associated with the Type II assessment program do occasion costs for district officials. They directly limit these officials' ability to do things like invest in other reforms or provide a savings to taxpayers.

The key point is that the locus of costs has important implications for the accounting of costs as well as for the behavioral responses to innovation. I shall pay explicit attention to the imposition of costs across categories of actors

in the trial cost analyses which follow in Chapters 3-6.

E. Discerning the Unit of Analysis

The results of cost comparisons of alternative approaches can be quite sensitive to the scale of the respective enterprises (Levin 1983). It can matter whether the comparison is between traditional assessment and an alternative approach within a school district, region, state, or nation. Scale economies can be important, and an analyst might find a small scale application of a reform is considerably more costly on a per unit basis than is a much larger undertaking.

In the empirical analyses of performance assessment costs which follow, I place primary emphasis on the individual state as the appropriate unit of analysis and address scale issues by providing cost estimates for hypothetical small, middle sized, and large states. However, I also treat certain development costs as more national in nature and apportion these costs across the participating states. This apportionment requires assumptions about how widely accepted performance assessment becomes as an education reform.

Care needs to be exercised when relying so heavily on relatively large units of analysis. One problem stems from the potential for aggregated data to gloss over sources of cost that are important at more micro-levels.

For example, the amount of time needed to train teachers as

scorers of performance assessments may vary substantially across LEA's, depending on things like the average amount of subject matter preparation present within a school district's faculty. At the state level, the localities requiring more resources for staff development will, to some degree, be balanced by those requiring fewer resources, but costs could vary substantially across local sites. Moreover, to the degree that large units like states vary in the incidence of difficult as well as easy to train teachers, there could be variation in costs across states as well.

In the cost analyses which follow, I deal with variability in how difficult it is to train professional staff by sketching alternative scenarios where there are differences in the average amount of training that is required. But this only begins to address the issue of variability across individual sites in the costs of implementing so sweeping a reform as the transformation of student assessment.

A related question about the relevant unit of analysis grows out of the realization that a reform as complex as the introduction of performance assessment techniques is not a monolith and contains any number of distinct parts. For example, the NSP proposal discusses alternative means by which assessment tasks will be developed. Some tasks will be developed internally by teachers and others working on the project; others will come from external sources and will need to be certified as meeting requirements set forth by the project leaders. The costs of developing assessment tasks can vary depending on the method

employed. For the sake of keeping my cost analysis tractable, I will make explicit assumptions about the origins and unit prices of tasks. These assumptions will be based, to the extent possible, on actual experiences with the alternative means of producing performance tasks.

Finally, there is an important distinction to draw between the costs of developing a system and the costs of operating the system once it has been developed. In the case of performance assessment innovations, there are substantial start-up costs that involve constructing the assessment tasks, testing their validity, achieving the initial inter-rater reliability, and so forth. There are also important maintenance costs. In my cost analyses, I shall be attentive to both the development and maintenance phases of the performance assessment reform.

F. Discerning Instances of Diminishing Marginal Productivity

Economic research has generated a number of propositions about the behavior of production processes that have important implications for magnitudes of costs. For example, if the relevant production processes are beset with sharply diminishing marginal productivities of key educational inputs, unit costs may be elevated, perhaps substantially, as additional inputs are supplied. Alternatively, the production processes may be such that diminishing marginal productivities are neither widespread nor pronounced, in which case the upward pressures on unit costs

will be minimal as more inputs are provided.

The central point here can be illustrated by sketching two alternative scenarios of performance assessment in education: one is a high cost scenario and includes an emphasis on diminishing marginal productivities, the other is a corresponding low cost scenario.

High cost scenario. This is a world beset with diminishing marginal productivities. They affect teachers as well as students and occasion the following results:

(1) At any given moment there is wide variation in the ability of teachers to benefit from the inservice assessment training that is offered as part of the performance assessment reform. Some teachers benefit significantly and quickly; others not at all or minimally.

(2) The current cohort of teachers also varies widely in how able they are to implement the assessments that are developed.

(3) The teachers least able to benefit from the available training are the teachers performing at the lowest levels.

(4) For all the teachers who are able to benefit from the available training, the magnitude of the gain in performance drops as they reach higher levels of performance.

(5) A similar set of phenomena arises with respect to students. Namely, students vary in their ability to benefit from the feedback provided by performance assessment; they vary in their level of performance; the lowest performing students are the least able to benefit from the feedback; and the marginal effectiveness of the assessment information drops off sharply (for all students) as they reach higher levels of performance.

If this portrait comes close to describing the real world of performance assessment, the cost of the enterprise will be very high. Such high costs may still be worth bearing, but it is clear

that their magnitudes will be substantial.

Low cost scenario. Here diminishing marginal productivities may be present, but their impact is much more modest. In this scenario, education is viewed as a cumulative process such that useful assessment information provided today makes learning tomorrow less costly. Moreover, the assumption is that there are important scale economies that are possible such that assessment tasks developed by teachers in one locale are readily transferable to others. It can be further assumed that as teachers gain experience at both developing and utilizing assessment tasks, it becomes easier to make effective use of performance assessment within classrooms. Finally, the assumption is made that assessment becomes so closely aligned with instruction that it no longer makes sense to conceive of it as a separate entity.

This is clearly a low cost scenario. If it is coupled with even conservative estimates of the potential benefits associated with the reform, the stage is set for finding a very favorable level of benefits in relation to costs.

Both the high and the low cost scenarios are plausible, but they both cannot be correct. Questions about which scenario is more accurate are ultimately empirical questions. However, the requisite empirical analyses will not be straightforward because proponents of performance assessment reform can easily claim that the high cost scenario, to the degree that it is played out as the reform is pursued, is more related to a failure to implement the

reform properly than it is to more fundamental flaws in the more intrinsic merits of performance assessment as a reform.

In the cost analyses which follow, I make extensive use of different scenarios to generate estimates of the costs that arise in alternative states of the world. I leave it to the reader to choose which of the scenarios (or which set of scenarios) seem most plausible.

Summary

In this chapter, I have explored a series of conceptual issues that are central to any attempt to estimate the costs of an educational innovation such as pupil performance assessment. For many of the problems there are no straightforward solutions, and my response is to proceed by conducting cost analyses for a series of three different sized states under a wide variety of assumptions. These assumptions will come in three varieties: best-case, middle-case, and worst-case, from the perspective of proponents of performance assessment reforms (i.e., the best-case is the case with the lowest cost estimates). In the final chapter, I report my results by gathering my estimates together under these headings. This has the effect of accentuating the differences between the best and worst case views since the scenarios build on one another in an exponential fashion. However, this is not necessarily a drawback, since one of my goals is to place upper and lower bound limits on the cost estimates.

Moreover, readers are certainly free to adjust the combinations of scenarios to more closely approximate their perceptions of reality.

CHAPTER 3

DEVELOPMENT COSTS

I. Introduction

The cost estimates I provide in this and the next 3 chapters are based on a conception of performance assessment and its role in promoting systemic reform that closely parallels the New Standards Project's proposal (NSP 1992). The focus will be on pupil performance assessment in two areas (mathematics and language arts), and at 3 grade levels (4, 8, and 10). The NSP is more ambitious than this and includes a commitment to the development of performance assessment of science and work readiness skills, but there is less information available here.

Much of what I will be calling development involves the production and refinement of tasks that serve as the basis of the pupil performance assessment system. The tasks that are produced enter a common bank from which participating school systems can draw. The results of the assessments are used in a variety of ways. They may enter individual pupil portfolios and complement the results of projects developed locally. They may also play a role in assessing the performance of teachers and/or entire schools. In any case, the assessment tasks play a central role, and their development can be thought of accurately as an investment activity. Once the task bank is in place, it can be

drawn upon over a period of time.

I assume that most of the development efforts take place during the first four years of the project, but I will also recognize a continuing need to develop new assessment tasks. The costs of these Year 5 and beyond development costs belong here rather than in the operations cost chapters because they represent investments that all participating school systems can draw upon.

It is worth noting that I have not included adjustments for changes in the price level over time in the following cost estimates. Instead, I am estimating future costs in terms of 1993 dollars. Adjustments for time preference will need to be made if there is interest in summing costs over time. I have also not dealt with differences in the costs of schooling inputs across states. A large literature has developed around this topic in recent years, and the interested reader can use the available indices as a basis for an additional set of adjustments. Barro (1991) has reviewed this literature and provides an overview of the available indices.

II. Development Cost Estimates For Years 1-4

I have divided the Year 1-4 development costs into the following categories: (1) administrative overhead; (2) production of usable assessment tasks; (3) initial task refinement; (4) production and distribution of pilot tests; (5) administration of the pilot test; (6) pilot test calibration; (7) scoring; and (8)

pilot test interpretation.

(1) Administrative Overhead

The production, refinement, pilot testing, and ultimate distribution of performance assessment tasks involves a considerable level of central administrative support. The production groups must be coordinated, materials need to be collected, disagreements resolved, and so forth.

I have based the following estimates of these costs on discussions with NSP staff about the level of administrative support that has been part of the Task Development process.

Annual Central Administration for Task Development

Professional staff	
.75 FTE @ 55,000	41,250.
fringe @ 33%	13,613.
Clerical support	
.50 FTE @ 20,000	10,000.
fringe @ 33%	3,333.

Space	
.33 of NSP space figured on a total monthly rental of $1800.	7,128.
Misc	
telephone, paper supplies, photocopying, postage, etc. (.33 of the NSP budget for these items)	5,000.
Subtotal Administrative Overhead	80,324.

(2) Production of Usable Assessment Tasks

I seek to develop "benchmark" figures for the average cost of developing a successful task that can be entered into a national Task Bank. The NSP has identified two primary means by which performance assessment tasks will be produced. The project has also generated unit cost estimates for each means. I have modified these unit cost estimates to reflect alternative

assumptions about how tasks ultimately will be produced over the period of development. Some of these assumptions reflect more favorably on NSP than do others. My hope is to provide a balanced view of what costs are likely to entail under alternative but always plausible conditions.

The modifications I make vary along three dimensions: (1) the degree to which one production method is used relative to another; (2) the degree to which productivity, regardless of the method chosen, changes over time; (3) the degree to which production costs, again regardless of the method chosen, vary across subjects.

In what follows I say more about each of these three modifications and establish the magnitudes that give rise to best, middle, and worst-case cost scenarios.

Productivity differences across alternative modes of production. The NSP will be generating new tasks from several sources. One of the most important of these is a broad network of practicing (front-line) teachers working together in groups to produce tasks. The project seeks to involve these front-line teachers in every substantive aspect of performance assessment, and the presumption is that "regular" teachers need to play a central role in the development of tasks. Enhanced credibility for the performance assessment program is one of the expected benefits of deeply involving teachers in the program. I shall call this a "generalized" mode of production.

The second source of new tasks relies less heavily on front-

line novices and more heavily on those who have already demonstrated their ability to produce tasks efficiently. I shall call this a "specialized" mode of production. Its origins lie in the early experiences of the NSP where it soon became apparent that some individuals were more fertile sources of good task ideas than were others. The NSP's early experiences also suggest that the ability of individuals to produce good tasks increases with practice. The NSP has responded to these early results by anticipating the use of more specialized, almost professionalized, modes of task production. These more selective and specialized sources of tasks may involve individual entrepreneurs (including teachers who may have been introduced to task generation through participation in the other production mode), textbook companies who set themselves up to produce performance tasks, and/or other types of vendors.

It is worth noting that the unit cost of producing a task may or may not be lowered through the use of these more specialized modes of production. Much depends on the size of the immediate as well as longer term supply of these more productive individuals. If the wages required to attract these individuals into task production are sufficiently high, the productivity gains that are generated may be more than offset. Much also depends on the actual returns to training (experience) and specialization. It may be the case that there are limits to how many good tasks individuals (or even teams of individuals) can produce. Practice may help at the outset, but once the individual or team reaches a

certain limit, productivity may drop off sharply.

Under the assumption that the productivity gains are real and long lasting and that the extra wage required to hire these individuals is lower than the dollar value of their additional productivity, the unit cost of producing tasks with the more capable and/or highly trained (experienced) manpower will be lower.

For our purposes here, I will assume that the unit costs of the more specialized method of production are lower. Later, I will attend to the problem of assigning dollar magnitudes to the respective unit costs.[1]

I shall also assume that for all three scenarios (worst, middle, and best-case) there will be an evolution during the development phase of the project away from the generalized and toward the specialized and, by assumption, less costly method of production. However, I shall impose a limit on the degree to which this shift takes place out of deference to the NSP's emphasis on keeping front-line teachers directly involved. Moreover, the ability even at the outset to rely on the less costly production method shall depend on the nature of the scenario. For example, I shall assume that for the worst-case scenario it is relatively difficult to make use of the less costly

---

[1] As NSP gains more experience with the development of tasks, there is a growing sense that the specialized mode of production will be relied upon more heavily than initially anticipated. This shift in thinking suggests that the unit costs associated with the specialized mode in fact are lower, at least in the short term.

method. Either credibility of the items in the field is compromised when so-called specialists' services are enlisted, or the cost of hiring these specialists begins to outstrip the savings that were anticipated. In either case, under the terms of the worst-case scenario, production will remain heavily dependent on the more costly generalized method.

In contrast, according to the best-case scenario, I shall assume that it is relatively easy to make the shift to the less costly method, even at the outset of the development period. It may turn out that teachers engaged early in the front-line teacher production method quickly gain the requisite skills and quickly enhance the supply of the specialized individuals who are, by assumption, more highly productive. This keeps the necessary wage premiums from rising to any great degree and makes it possible to realize the savings. In addition, it may be the case that relatively low levels of direct involvement of front-line teachers are necessary for the system to have credibility in the field. The middle-case scenario deals with a reality lying between these two more extreme views.

To be specific, I assume for the best-case scenario that the production ratio begins at 80/20, meaning that 80% of the tasks are produced using the more costly front-line teacher method while 20% are produced according to the less costly more specialized method, and reaches 20/80 by year 4. For the middle-case scenario the ratio begins at 90/10 and reaches 60/40 by year 4. For the worst-case scenario, the ratio begins at 100/0 and works its way

to 70/30 by year 4. Table 3-1 provides a summary of my assumptions about how the mix of production modes will vary over time.

Table 3-1: Assumptions Regarding Changes in the Mix of Task Production Modes Over Time

	Year 1	Year 2	Year 3	Year 4
Best	80/20	60/[illegible]	40/60	20/80
Middle	90/10	80/2[illegible]	70/30	60/40
Worst	100/0	90/10	80/20	70/30

(The first number in each cell represents the percentage of tasks coming from the generalized method; the second number represents the percentage share coming from the specialized method.)

Productivity gains over time. The second modification involves a series of assumptions about the rate at which the productivity of task generators (both those working within the generalist and the specialist frameworks) improves with time. This leads to a series of assumptions about the degree to which raw tasks survive pilot testing and the other reviews and lead to usable tasks that ultimately enter the task bank.

The benchmark I use here is the early experience the NSP had with the production of tasks. Early task production within the project was carried out by teachers working together during large national meetings. Experience demonstrated that on average roughly 50% of these earliest tasks survived the subsequent

reviews and pilot testing. I used this 50% figure as the starting point for all three scenarios on the grounds that it is the best available indicator of the kind of production difficulties that need to be overcome. Where the worst, middle, and best-case scenarios differ is in the treatment of how quickly the production difficulties are overcome. I further assume that the underlying learning curve is such that the initial improvement (from year 1 to year 2) is larger than subsequent improvements. Table 3-2 describes these assumptions for each scenario and year.

Table 3-2

Year to Year Improvements in Productivity

	Year 1	Year 2	Year 3	Year 4
Best	50	75	85	90
Middle	50	70	75	80
Worst	50	60	65	65

Cell entries reflect the percentage of raw items that survive subsequent refinement and pilot testing.

Productivity differences across subjects. One of the interesting results of the early work on task development within the NSP was a clear difference across subject matters in the rate

at which successful tasks were developed. In particular, more successful mathematics tasks were produced relative to language arts tasks following comparable development efforts. There are several possible reasons for the discrepancy--perhaps the most compelling is the claim that more progress has been made within mathematics to reach consensus and clarity over the skills and capabilities that the nation is seeking to convey to its students. If this is the reason and if greater clarity in the language arts area is developing,[2] then we can expect the additional cost NSP encountered in the early development of language arts tasks to decline with time. If, on the other hand, the discrepancy is due to intrinsic differences across the curricula, or if it proves to be impossible to reach consensus over curriculum content within language arts, the differential is likely to persist and may even widen.

There is a related question to ask about whether the costs of developing consensus regarding curricular content are properly charged to the performance assessment enterprise. Recall from Chapter 2 that this is a locus of cost issue. To the extent that greater curricular clarity is an important education reform irrespective of its effects on assessment, a case can be made for at least pro-rating the cost of achieving greater clarity across its many applications.

---

[2] The New Standards Project and the U.S. Department of Education are currently sponsoring jointly an initiative to foster the development of curriculum content standards for language arts.

I have handled these interrelated issues by exploring the cost implications of alternative assumptions about the sources of the observed differences in task development across subjects and across time.

In particular, I assume that there are differences across subject areas in the level of curricular clarity and that relatively high levels of ambiguity make it more difficult to generate performance assessment tasks. Moreover, I assume that the efforts being made to foster curricular clarity have significant but variable payoffs (where the magnitude of the payoff depends on how optimistic the scenario is) over time in terms of reducing the difficulties associated with producing assessment tasks. In other words, I am assuming that while efforts to foster curricular clarity will make it easier to generate tasks, there will remain differences (except in the case of the most optimistic, best-case scenario) across subject areas in how much difficulty is associated with task development. In the corresponding worst-case scenario, I assume that the differences are due entirely to the intrinsic natures of the two subject areas and that efforts to generate agreement about what to teach has no impact on how difficult it is to produce performance task items.

Finally, I assume that the costs of fostering curricular clarity are not reasonably charged to performance assessment. Thus, I allow some but not all of the costs associated with curricular ambiguity to be associated with performance assessment.

The details of these assumptions are spelled-out in Table 3-3.

Table 3-3

	Year 1	Year 2	Year 3	Year 4
Best	2:1	1.6:1	1.3:1	1:1
Middle	2:1	1.8:1	1.6:1	1.4:1
Worst	2:1	2:1	2:1	2:1

Each cell entry depicts the number of mathematics usable mathematics tasks produced for each usable language arts task.

Unit Cost Valuation and Levels of Production

The NSP has projected task development costs into the future and these projections provide the starting points for the cost analysis. In particular, the NSP estimates that a raw performance assessment task can be produced, on average, using the generalized mode of production for $2,000 and that the comparable figure for the specialized mode is $1,000.[3]

[3] These unit costs per raw task were calculated from unit estimates provided by the NSP for usable tasks. Specifically, NSP estimates that it costs $4,000 per usable task produced using the generalized method. In year 1, the loss of tasks under all three scenarios is 50%. This suggests a raw task unit cost of $2,000.

I have also assumed that for each subject and grade level, the goal is to produce 25 usable tasks per year over the 4 year development period. Production at this rate will yield a Task Bank of 100 units per subject per grade level at the end of the development phase of the project.

Table 3-4 describes the costs of generating these 100 usable tasks for each combination of grade level and subject area, and takes into account the adjustments described above.

Table 3-4 About Here

Notice that I have reflected some of the costs of task refinement in terms of the distinction between raw and usable tasks. In addition, there are the direct costs associated with accomplishing the refinement. It is to these direct costs of task refinement that I turn next.

(3) Initial Task Refinement

The production of raw tasks is followed immediately by an initial review by subject matter and measurement specialists. The NSP has established two centers which coordinate this work, one for mathematics and one for language arts. The review consists of

---

The corresponding estimate for usable tasks under the specialized method is $2,000. This translates into a raw task unit cost of $1,000.

an informal pre-pilot which generates initial feedback from the field as well as more systematic reviews of how well aligned the tasks are to existing curriculum frameworks.

The NSP financial planning documents reveal the level of spending on these activities and include projections over the development phase of the project. I rely upon these estimates as the basis of my cost calculations for this aspect of the review process. It does not appear necessary here to depict alternative scenarios.

The estimated expenditure for the initial refinement of mathematics tasks is

Mathematics

Production Staff	250,000
Outside Consultants	25,000
Advisory Committee Meetings	50,000

I shall assume the same costs for Language Arts.

Language Arts

Production Staff	250,000
Outside Consultants	25,000
Advisory Committee Meetings	50,000

The total cost is: $650,000.

(4) Production and Distribution of Pilot Tests

The NSP has projected costs on the order of $300,000 for the printing and distribution of the exams that will be used in a pilot test program covering 2 subject areas but only 2 grade levels. In order to make this figure comparable to the 3 grade level prototype being considered here, it is necessary to make a 50% adjustment upwards (the underlying assumption being that the production and distribution costs are evenly divided across the grade levels).

The adjusted figure is, therefore, $450,000.

(5) Pilot Testing

The formal pilot testing for the raw tasks which survive the initial review involves a selection of 10 schools from each of the 22 partners participating in the NSP. I shall treat this as an appropriate number of sites for purposes of establishing the necessary levels of reliability and validity before a task can enter the project's task bank.

Within each selected school, three classes per subject per grade level participate in the pilot test. This yields a total of 660 classrooms for each subject and grade level. I assume that 2 of the raw tasks are administered within each participating classroom, and that the 2 tasks require 6 hours of class time plus 2 hours of prior teacher preparation. My cost estimates here are based on the cost of the projected amount of teachers' time that

is involved.[4]

Cost of Teacher Time	
660 * 8 * $25	132,000
Total	132,000.

The $132,000 pertains to one subject and one grade level. While there may be some economies of scale to be realized in the conduct of these pilot studies, I will assume these are negligible and raise the $132,000 figure by a factor of 6 to account for the costs of pilot tests for 2 subjects at 3 grade levels.

Total for 2 subjects at 3 levels	792,000.

Assuming that the classes in which these pilot tests are administered hold an average of 25 students, the total number of tasks that can be graded following the pilot tests is:

660 * 2 * 25 = 33,000 tasks per subject and per grade level

This yields a total of 33,000 * 6 = 198,000 tasks in total that are available for scoring following the pilot test.

---

[4] While there are additional costs to consider (e.g., the costs of students' time, administrators, space, etc.), they are either not easily expressed in a common metric (e.g., the cost of students' time) or they are of small magnitude. I have excluded these costs from these analyses.

(6) Pilot Test Calibration

Pilot test results are used as the basis of developing rubrics and benchmarks for each of the identified tasks. The goal is to achieve clarity in the ranking of different possible responses to the tasks. The NSP experience suggests that these benchmarks and rubrics are most efficiently established by convening a national meeting involving approximately 80 specialists. Samples of the pilot test results are used to calibrate the rubrics that are established. The 80 specialists deal with both subject areas and all three grade levels.

The estimated costs of these meetings, including honoraria for the participants, is on the order of $25,000 per subject per grade level. Thus, for a testing program that involves 2 subjects and 3 grade levels, the annual cost will be on the order of $150,000.

(7) Scoring the Pilot Tests

For the sake of keeping the analysis tractable, I shall envision two levels of involvement for teachers. The first level consists of teachers who participate in the initial scoring of the pilot tests. Based on the NSP experiences, the training of teachers and others in the use of performance assessment is closely linked to scoring practice. The teachers that do the initial scoring of the pilot tests will acquire a sophisticated

understanding of performance assessment and are expected to play a central role in the subsequent training of the second tier of front-line teachers who will be directly involved in the administration and scoring of tasks when the system is fully operational.

The NSP early experiences with task scoring suggest that a well trained teacher can be expected to score 10 tasks per hour, on average, or roughly 50 tasks per day. The NSP estimates that teachers will be able to reach this level of productivity as scorers following two days of supervised scoring.

It has been the NSP practice to identify 2, 3, or 4 teachers from each of the 22 partners for each of the grade levels and subjects to participate in the scoring of pilot exams. These individuals come together for a national 5 day meeting, and their assignment is to score between 20 and 30 per cent of the 198,000 tasks that were generated during the pilot testing.

The estimated costs of a 5 day national meeting for the roughly 400 people that will attend this meeting (assuming, on average 3 teachers for each subject and grade level from each of 22 partners) are as follows:

Cost of a 1 Week National Training Program for 400 participants

400 honoraria figured at $100 per day	
400 * 100 * 5	200,000

travel $800 average	320,000
lodging $80/night (5 nights)	160,000
food and misc $40 per 400 for 5 days	80,000
fees and expenses for instructors assuming 4 instructors working full time with every 100 participants	
16 instructors at $250/day for 5 days	20,000
travel for 16 instructors at $800	12,800
lodging at $80 for 16 instructors for 5 days	6,400
food and misc $40 per 16 instructors for 5 days	3,200
Total 1 Week Training Program	$802,400

The figures I am using for per diem stipends warrant comment. I am basing them on the NSP practice of paying flat $100/day fees to front-line teachers who participate in the program. Once a teacher has become knowledgeable about performance assessment and is in a position to provide instruction to others, I raise the stipend to the $250/day level. These figures are in line with current NSP practice.

On the assumption that these teachers will score the 50 tasks per day, 400 * 50 * 5 = 100,000 tasks out of the possible 198,000 tasks will be scored. However, since the 50 task per day rate is what NSP finds trained teachers can accomplish, an adjustment downward needs to be made to account for the fact that at the beginning of the week the teachers will not be fully trained.

My assumption is that the number of tasks scored each day will vary according to the following schedule:

M	T	W	TH	F
5K	10K	20K	20K	20K

(The assumption underlying this schedule is that a fully trained teacher can score 10 tasks per hour for a 5 hour day, and that it takes two days to reach this level of proficiency. On day 1, I am assuming the average rate is 2.5 tasks per hour; on day 2, my assumption is that the average rate is 5 tasks per hour.

Under these assumptions, the week generates a total number of 75,000 scored tasks. This leaves 123,000 tasks that need to be scored at the local level.[5]

I assume further that the local scoring will be accomplished through a series of regional meetings that take place within each of the partners in the project. Recall that 10 schools were

[5] I am assuming that the tasks used to create benchmarks and rubrics are not removed from the pool of tasks that need to be scored.

identified within each partner for participation in the pilot test. I shall assume that there will be one regional meeting for each group of 10 schools, and I shall calculate the number of participants that will be needed at these meetings on the basis of the number of pilot test tasks that need to be scored. Finally, I shall assume that the same learning curve applies to the local scorers as applied to the teachers at the national meetings.

There is a trade-off between the number of teachers involved in the scoring process and the amount of time each teacher is expected to devote to scoring tasks. On the assumption that the NSP seeks wider involvement of teachers, I shall limit to 4 days the amount of time any one teacher devotes to scoring. And I shall assume that each regional scoring (training) meeting will involve no more than 30 teachers. (These assumptions correspond to NSP practices.)

Cost of a 30 person 4 day regional scoring/training meeting

honoraria for 30 at $100/day for 4 days	12,000
ground travel ($10 per day/per participant)	1,200
lodging (n.a. for participants)	
materials ($10/participant)	300
meals ($10/participant/day--lunch only)	1,200
space	n.a.
leaders (assuming 2 at $250/day)	2,000
leaders prep (assuming 2 days at $250)	1,000

lodging for leaders (80 * 2 * 4)	640
meals for leaders (40 * 2 * 4)	320
travel for leaders (estimated)	150
	18,810

Each regional meeting yields 30 trained scorers plus 4,125 scored tasks (assuming the same learning curve for scorers that I used above).

I am assuming the goal is to score all of the remaining pilot test results. There are several possible reasons for this. For example, results from the full sample may be necessary for psychometric reasons. Also, more scoring offers an opportunity to create a larger network of trained teachers, particularly if the regional workshops are structured so that new teachers are invited to participate.

For now, I provide estimates of the costs associated with scoring all of the pilot test results. Later, when attention turns to the operations phase of the project, I will deal with questions about how many front-line teachers and trained local teachers are needed.

Recall that there remain 123,000 tasks that need to be scored. Each regional workshop generates 30 trained front-line teachers and 4,125 scored tasks. Thus, the number of regional workshops required is 123,000 / 4,125 or 29.8 and I will call this 30. If each regional workshop costs $18,810, the total cost of

the local scoring program will be 18,810 * 30 or $564,300.

Thus, the total costs of scoring are:

$802,400 + $564,300 = $1,366,700.

One of the by-products of this expenditure will be a reservoir of 30 * 30 or 900 trained front-line teachers (assuming there are no repeating teachers in the regional meetings). These may or may not be evenly distributed across the participating states or units, depending decisions about where the meetings are held. I will deal with these distributional issues later in the operations sections of this analysis (see Chapters 4-6).

(8) Pilot Test Interpretation

The NSP has budgeted roughly $30,000 to cover the costs of interpreting the results of the pilot tests in two subject areas over 2 grade levels. I shall use this figure as the basis of my cost estimate, but I will raise it by 50% to allow for the third grade level that is envisioned in this analysis.

Thus, the baseline figure for pilot test review and interpretation will be $45,000.

III. Year 5 and Beyond Development Costs

Continued Task Development

If the goal is to establish a working bank of 100 tasks, and if the shelf-life of each task is assumed to be 10 years, and if the development phase of the project lasts 4 years and generates 25 new tasks each year, it is easy to show that the production of 10 new tasks per year during the operations phase of the program will bring the system into equilibrium after 14 years. The key point here is that there will be some number of new tasks that needs to be produced each year during the operations phase of the program. For the purpose of this cost analysis, I will assume that the goal will be to produce 10 usable new tasks each year beginning in Year 5.

Recall the alternative production modes through which tasks can be produced, each with its own implications for costs. I shall retain the distinction among best, middle, and worst case scenarios, and I shall carry forward the assumptions I made about the mix of alternative production modes. In particular, I shall use the year 4 mix and assume that no further changes are made over time. Recall that the best-case scenario involves a 20/80 mix of generalized and specialized production modes. The corresponding mixes for the middle and worst-case scenarios are 60/40 and 70/30, respectively.

For simplicity's sake, I will assume that the productivity gains that were realized during the development phase of the project flatten and that there are no further gains to consider.

Again, there are differences among the best, middle, and worst-case scenarios, and I will carry forward the percentage yield figures that I used for Year 4 of the Development Phase (see Table 3-2). Accordingly, the best-case yield is 90%, the middle-case is 80%, and the worst-case is 65%.

I shall also carry forward the Year 4 assumptions about productivity differences across the two subjects (see Table 3-3). Under the terms of the best-case scenario, there is no difference. Under the terms of the middle-case scenario, the difference is on the order of 1.4:1.0. The corresponding figure for the worst-case scenario is 2:1. I will assume that these ratios remain fixed throughout the Operations Phase of the project.

Table 6 provides the overview of the costs of producing 10 usable tasks per year beginning in year 5 of the project (Year 1 of operations).

Table 3-5

Costs Associated with Continuing Task Development

In Years 5 and Beyond

Number of Required Raw Tasks for Mathematics	
Best	11.1
Middle	12.5
Worst	15.4
Average Unit Cost	
Best	$1,200
Middle	$1,600

Worst	$1,700
Total Costs for Math	
Best	$13,320
Middle	$20,000
Worst	$26,180
Total Costs for Language Arts	
Best	$13,320
Middle	$28,000
Worst	$52,360
Total Costs for Math + Language Arts	
Best	$26,640
Middle	$48,000
Worst	$78,540

These continued task development costs will need to be shared across the various participating states and units. In this sense they can be thought of as being developmental, but they occur during the operations phase of the project. Because these activities will take place in the context of on-going operations, I will assume that the associated administrative costs including the costs of pilot testing, task calibration, scoring, and the like will be absorbed within the operations costs that are described in the following chapters.

IV. Summary

Table 3-6 provides an overview of the Year 1-4 plus the Year 5 and Beyond Development Costs that have been identified.

Table 3-6 About Here

According to the Table, development costs will range between $4.34 and $4.43 million dollars in Year 1. These costs drop and the difference between the best and worst case scenarios widens over time, so that by the time Year 4 arrives, the low estimate is $3.73 and the high estimate is $4.12 million dollars. To put these figures in context, they can be expressed on a per pupil basis. If we reason that the tasks and teacher skill levels that are developed during this period are available to all pupils in the 17 NSP participating states, the total pupil population being served is on the order of 18.905 million. This corresponds to a per pupil cost of between 23.0 and 23.5 cents in Year 1. By Year 4, these figures drop to 19.8 and 21.8 cents, respectively (in 1993 dollars).

The following three chapters shift the discussion to the costs of operations in three different sized prototypical states.

CHAPTER 4

OPERATIONS COSTS IN A LARGE STATE

I. Introduction

I consider Operations Costs from the perspective of three states: large, mid-sized, and small, and I devote a chapter to each. In Chapter 7, I contrast the results and comment on the role played by economies of scale. The focus in this chapter is on a large State where I assume there are 4,100 elementary schools and 1,381 secondary schools organized into 1,000 local education agencies (i.e., school districts). I will assume further that the state's grade 4 enrollment is 255,832 pupils and that the enrollments in grades 8 and 10 are 237,387 and 223,162, respectively.

I deal explicitly with the following components of operations costs: (1) Supplemental Lead Teacher Training; (2) Scorer Training; (3) Continuing Scorer Training; (4) Outside Auditing; (5) Administration of Tasks; (6) Scoring; (7) Utilization of Results; and (8) Administration and Overhead (including the costs of printing and distributing the exams). Next, I consider alternative assumptions regarding the possible absorptions of selected cost components. The chapter concludes with an overview

and summary of the cost estimates.

II. Components of Cost

(1) Supplemental Lead Teacher Training

Available Supply of Lead Teachers

Recall that as a by-product of the Development Phase there is a pool of trained scorers. Under the assumptions I imposed in Chapter 3, I estimated the size of this pool per year to be 1,300 trained scorers per year (400 trained at the national scoring meetings and 900 trained regionally). At the end of the 4 year Development Phase, the maximum number of trained scorers will be 1,300 * 4 or 5,200, assuming there are no scorers who repeat their training program. This also presumes that there is no loss of skill over as long as 4 years for teachers who learn to score at the outset of the project.

Given the likelihood that scorers will vary in how well they learn the requisite skills, that some decay will take place over time for those who are trained early, and that the project will lose track of some participants, I will make alternative assumptions regarding the actual size of the reservoir of scorers that is available at the end of the Operations Phase.

According to the best-case scenario, there is little loss over time and teachers learn the relevant skills quite easily and

uniformly. In other words, NSP does not have to deal with significant unevenness in how well teachers learn to be scorers. Nor is there much unevenness in how well the trained teachers retain their skills. Nor does the project lose track of many scorers over time. The middle and worst-case scenarios relax these assumptions and introduce potentially significant levels of unevenness, depreciation and obsolescence, and loss.

There are no obvious benchmarks to rely upon in assigning magnitudes to the discount factors that need to be used, so I will make the relatively arbitrary assumption that under the best-case scenario, the effective loss is 10%, and that under the middle and worst-case scenarios, the respective percentage losses are 20 and 30.

My assumption is that these experienced scorers constitute the initial NSP representation in the field. These people will play lead roles in the training and implementation of the project within the participating states. They will be involved in both the performance assessment as well as the cumulative portfolio development aspects of the NSP. I will refer to them as Lead Teachers.[1]

I also assume that these Lead Teachers are divided across the

---

[1] According to NSP documentation, performance tasks constitute just one part of the cumulative portfolio that will be generated for each student. While there will be central guidance provided about the types of items that should be included in students' portfolios, much discretion will be maintained at the individual school and teacher levels. The Lead Teachers will provide training and assistance to front-line teachers who are participating in the project.

participating states in proportion to the respective states' populations. My rationale for this is based on the NSP practice of varying the number of invitations to the national scoring meetings according to its partners' populations (recall that either 2, 3, or 4 teachers from each grade level and subject were invited).

Recall that this state is relatively large with 4,100 elementary schools and 1,381 secondary schools with a total pupil enrollment in grades 4, 8, and 10 of 716,381. I assume that this state received 4 * 2 * 3 or 24 nationally trained scorers each year (during the Development Phase), and that the number of regional training workshops that were conducted within the state is proportional to the state's share of the NSP base student population (i.e., the population from all the participating states and units). According to NSP documentation, a state with 716,381 pupils would comprise 14.8 per cent of the pupil base being served by the project. Thus, I assume that it operated 14.8 per cent of the 30 regional workshops that were held each year. This corresponds to 4.44 workshops per year. Recall that each workshop generated 30 trained scorers. It follows that for each year the large State created a pool of 4.44 * 30 or 133.2 locally trained scorers. This yields a total of 24 + 133.2 or 157.2 potential Lead Teachers each year for a total possible of 628.8. The application of the best, middle, and worst case loss rates that I derived above generates the following estimates of Lead Teacher Supply for the large State at the close of the Development Phase

of the Project:

Best Case

$$628.8 - (.1 * 628.8) = 565.92$$

Middle Case

$$628.8 - (.2 * 628.8) = 503.04$$

Worst Case

$$628.8 - (.3 * 628.8) = 440.16$$

Demand for Lead Teachers During Operations

The next question is whether the supply of these Lead Teachers thanks to the Development Phase of the Project will be adequate to staff the Operations Phase of the Project. To begin to answer this question, I make a series of assumptions about the scope of the operational phase of the performance assessment project. There are two dimensions to this demand: (1) the number of schools that will be involved in the operational version of performance assessment; and (2) the level of direct supervision by Lead Teachers that is required within each participating school. I will be making alternative assumptions regarding each dimension, and I shall join them within the scenario framework so that the best case of one is linked with the best case of the other. In other words, I will not be considering alternative combinations of best, middle, and worst cases along each dimension.

Counts of Participating Schools

At one extreme, I will assume that in order for the project to achieve its goals, it will be necessary to implement an annual performance assessment program within every school in each participating state. I call this a census approach to implementation, and it corresponds to a worst-case scenario with respect to the associated costs.

Recall that a major goal of the NSP is to change fundamentally the conduct of instruction throughout entire schooling systems. According to this worst-case cost scenario, it is necessary to have an NSP presence within every school during every year of the operations phase of the project. I also make alternative assumptions about the level of the presence that is required, but for now the focus is on how many schools need to participate in a given year during the operational phase.

At the other extreme, I will assume that it is possible for the project to achieve its goals through the use of a light matrix sampling design. The presumption here will be that a periodic program of assessment within a relatively small sample of schools is sufficient within each state to achieve the far-reaching goals of the NSP. The sample of schools and classrooms participating will vary from year to year. All schools and the relevant classrooms will be eligible for selection, and at any given time teachers and administrators will not know when their classrooms and schools will participate. Moreover, in any given year, I assume that the state will focus on some subset of the possible

tasks within the Task Bank. This scenario will correspond to a best-case view of costs since fewer resources will be required (by assumption) for the project to achieve its goals.

The middle case scenario involves a situation where there is interest in district specific results. In contrast to the census and matrix approaches, the presumption here is that there is interest in district level performance. The design will require sampling from within districts and this will require a measure of performance assessment that lies between the first two extremes that I have identified.

Level of Direct Supervision Provided By Lead Teachers

According to the NSP proposal, a goal of the project is to have two externally trained and certified scorers within each school participating in the performance assessment activities. I am assuming that such people correspond to what I have called Lead Teachers, and I note that there is some ambiguity surrounding the precise level of Lead Teacher supervision that will be appropriate. At one extreme, it could be that two Lead Teachers could handle all of the testing taking place within a school regardless of the subject being taught. Thus, in a secondary school with grades 8 and 10 present, two Lead Teachers could handle the testing program for both mathematics and language arts. From a cost perspective, this extreme corresponds to a best-case scenario.

At the opposite extreme, it may be necessary to have two Lead Teachers for each grade and subject being assessed. In this case, a secondary school with two grade levels would require 8 Lead Teachers. This reality corresponds to a worst-case scenario in terms of costs.

A middle ground can be defined by thinking of the Lead Teachers as being able to cross grade levels but not subject areas. Under the terms of this middle-case scenario, the secondary school with grades 8 and 10 would require 4 Lead Teachers.

These three scenarios (for both the number of participating schools and the number of needed Lead Teachers in each school) are used below to define the demand for Lead Teachers in a typical operational year of the project.

Best Case Scenario

Number of participating schools. This scenario involves the use of a matrix sampling design. I assume that the sampling goal will be 100 observations per task,[2] and that in any given year the State will employ 25 per cent of the tasks available within the Task Bank. If every student participating in the program received one task, implementation would require 25 * 100 or 2,500 pupils

[2] According to the NSP, a sample of 2,500 observations needs to be drawn for all 25 tasks for each subject at each grade level being considered to satisfy psychometric concerns over validity.

per grade per subject. But, I will also assume that each student participating will complete 2 tasks, and thereby reduce the required number of participating students by half to 1,250 per grade per subject.

Recall that the large State has a grade 4 enrollment of 255,832 and a population of 4,100 elementary schools. If the grade 4 students are evenly distributed across the schools, it follows that the average school will enroll approximately 62 4th grade students. If the goal is to have 2,500 participating 4th grade students (1,250 per subject), in a given year performance assessment will need to take place within approximately 40 of the 4,100 elementary schools (2,500/62).

At the secondary level, there are two grade levels. For the large State, there are 237,387 8th grade students and 223,162 10th grade students. There are 1,381 secondary schools and assuming all 8th grade students are enrolled in secondary schools and that the students are evenly distributed across the schools, it follows that each school enrolls, on average, approximately 165 students at each of these grade levels. If the goal is to have 2,500 participating 8th and 2,500 participating 10th grade students (again, 1,250 per subject), then in a given year performance assessments will need to take place within approximately 15 of the 1,381 secondary schools (2,500/165).

Level of staffing. In keeping with the best-case scenario, I assume that each participating school needs 2 Lead Teachers and that these Lead Teachers can handle both multiple subject areas

and grade levels (where they occur). If there are 40 elementary schools in the program, there will need to be (40*2) or 80 Lead Teachers for the elementary schools. If there are 15 secondary schools in the program, there will need to be (15*2) or 30 Lead Teachers for the secondary schools.

Thus, the best-case scenario involves a total annual demand of 80+30 = 110 Lead Teachers. This compares with the derived supply of 565.92. Thus, under terms of the best-case scenario, the large State will not need to provide supplemental training for Lead Teachers, at least not at the outset of operations. The costs of supplemental Lead Teacher training will be considered 0 for the best case scenario.

Middle Case Scenario

Number of participating schools. Here the idea is that the state is interested in having information from each district, and the presumption is that the matrix sampling design described above misses a significant number of districts. As I indicated earlier, the large State operates 1,000 separate school districts. I assume that the average grade 4 enrollment within each district is 255,832 / 1,000 or 256, and that the average grade 8 and 10 enrollments are 237,387 / 1000 or 237 and 223,162 / 1,000 or 223, respectively. Using the 62 4th grade pupils per school and 165 8th or 10th grade pupils per school figures that I derived above, it follows that on average each district operates 4.1 elementary

schools and 1.4 secondary schools.

I assume that a sample of 2 elementary schools per district and 1 secondary school per district will be adequate to provide the district level aggregates. This will require staffing performance assessment activities in 2,000 elementary schools and 1,000 secondary schools in a given year.

Level of staffing. In accordance with the middle-case scenario where the assumption is that 2 Lead Teachers are needed for each subject within each school, there is an implied demand 2,000 * 2 * 2 = 8,000 Lead Teachers for the elementary program and 1,000 * 2 * 2 = 4,000 Lead Teachers for the secondary program.[3] This means the state needs a pool of 12,000 Lead Teachers compared to the 503 that are available following the Development Phase.

Implications for costs. I assume the necessary training will take the form of a supplemental series of 4 day workshops structured around scoring exercises. The same costs that I derived earlier will apply. Recall that these workshops cost $18,810 and yielded 30 trained scorers. Thus, the supplemental cost for Lead Teacher training for the large State according to the middle case scenario will be:

((12,000 - 503) / 30) * $18,810 = $7,208,619

---

[3] Recall that the assumption is that Lead Teachers can cross grade levels but not subject areas. This explains why the secondary schools require 4 rather than 8 Lead Teachers.

Worst Case Scenario

Number of participating schools. Recall that the worst-case scenario involves a census approach to performance assessment where the presumption is that every 4th, 8th, and 10th grade student needs to be assessed every year in both subject areas.

The large State has a population of 4,100 elementary and 1,381 secondary schools. If the state pursues a census approach, Lead Teacher staffing will be required in each of these schools.

Level of staffing. According to the worst-case scenario, 2 Lead Teachers are needed for each possible combination of subject and grade level. Assuming elementary schools involve only grade 4, the total number of Lead Teachers needed for the 4th grade assessment program will be 4,100 * 2 * 2 or 16,400. The corresponding number of Lead Teachers for the 8th and 10th grade assessment programs (assuming they are all located within the secondary schools) will be 1,381 * 2 * 2 * 2 or 11,048.

Implications for costs. The total number of Lead Teachers needed according to this scenario is 16,400 + 11,048 = 27,448. In contrast, according to the worst-case scenario, the Development Phase of the project generates a supply of 440 Lead Teachers. The relevant cost calculation (assuming the Lead Teachers are trained through the use of regional workshops) is:

((27,448 - 440) / 30) * 18,810 = $16,934,016.

(2) Scorer Training

Best Case Scenario

Number of Scorers Needed

Assuming each participating student generates 2 tasks, the annual total number of tasks that need to be scored will be the number of students per grade level (2,500) * the number of grade levels (3) * the number of tasks completed (2) = 15,000.

I assume that each scorer scores 400 tasks. This is the equivalent of 8 days of work. The NSP does not seek to develop a supply of "professional" task scorers. It is, instead, committed to achieving a broad base of participation among teachers and others. For this reason, I impose the 400 task ceiling.

If there are 15,000 tasks to score in the large State, and if each scorer scores 400, the demand for scorers will be 37.5.

Level of Training Required

Minimal training will be required to train local scorers under terms of the best-case scenario. The underlying assumption is that this kind of assessment and its scoring will be very much in-line with how teachers think and go about their work. The

teachers are presumed to adapt quickly and easily. I assume that the training can be done quite informally within the local districts; as a consequence travel costs for participants become negligible and will be omitted. Since the Lead Teachers will be traveling, I have included an allowance for this travel in the budget.

However, for the sake of deriving cost estimates, I will continue to treat the training as if it has a group workshop nature. In particular, I will assume that what is necessary is the equivalent of a one-day workshop for 30 participants where the participant/Lead Teacher ratio is 8:1.

$100 per diem for 1 day for 30 participants	3,000.
travel:$20 average for 30 participants	n.a.
lodging:	n.a.
food and misc $10/day per participant	300.
materials (20/participant)	600.
Lead Teacher costs	
assuming 3.7 Lead Teachers per 30 participants	
$250 per diem per day per Lead Teacher	925.
travel costs at $40 per Lead Teacher	148.
lodging	n.a.
food and misc.: $10/day per instructor	37.
Total Cost per 1 Day Scoring Workshop	$5,010.

The yield for this workshop is 30 trained scorers. If the need for scorers is 37.5, the costs of training these individuals will be $6,263.

Middle Case Scenario

Number of Scorers Needed

According to this scenario, there will be assessment activities in 2,000 elementary schools and 1,000 high schools. The average number of 4th grade students per elementary school is 62; the average number of 8th and 10th grade students is 165. Thus, there are 454,000 students participating in a given year. If each student completes 4 tasks (two for each of two subjects), there will be 1,816,000 tasks to score.

If scorers score 400 tasks each, there will be a demand for 4,540 scorers.

Level of Training Required

A more ambitious level of training is required under the terms of the middle-case scenario. Instead of the equivalent of a 1 day (30 person) scoring workshop, I will assume that 2 days are

necessary. I shall also assume that a more intensive training experience is necessary. Instead of the 8:1 ratio of participants to Lead Teachers, I will assume that a 4:1 ratio is necessary. I shall also build travel costs into the budget, since my presumption is that it will be less possible for the training to place informally at the home sites.

$100 per diem for 2 days for 30 participants	6,000.
travel:$20 average for 30 participants (per day)	1,200.
lodging:	n.a.
food and misc $10/day per participant	600.
materials (20/participant)	600.
Lead Teacher costs	
assuming 7.5 Lead Teachers per 30 participants	
$250 per diem per day per Lead Teacher	3,750.
travel costs at $20 per day per instructor	300.
lodging	n.a.
food and misc.: $10/day per instructor	150.
Total Cost per 2 Day Scoring Workshop	$12,600.

Assuming there are 4,540 scorers that need to be trained, and assuming these 2 day workshops each yield 30 trained scorers, the cost of scorer training will be $1,906,800.

Worst Case Scenario

Number of Scorers Needed

The large State operates 4,100 elementary schools and 1,381 secondary schools. Assuming there are 62 4th grade students per elementary school and 165 8th and 10th grade students per secondary school, there will be 709,930 students being assessed each year. If each student completes 4 tasks, there will be 2,839,720 tasks to score each year.

Assuming each scorer scores 400 tasks, there will be a demand for 7,099 scorers.

Level of Training Required

Since this is the worst case scenario, I assume that teachers, on balance, find it difficult to grasp the requisite skills to function effectively as scorers. I assume that these teachers need to spend the equivalent of 4 one-day workshops acquiring these skills, and that these workshops will be offered regionally. I assume further that the scorer/Lead Teacher ratio in the workshop needs to be 2:1.

For each 4 one-day elementary task scoring workshop, there will be the following costs:

$100 per diem for 4 days for 30 participants	12,000.
travel:$20 average for 30 participants (per day)	2,400.
lodging:	n.a.
food and misc $10/day per participant	1,200.
materials (20/participant)	600.
Lead Teacher costs	
assuming 15 Lead Teachers per 30 participants	
$250 per diem per day per Lead Teacher	15,000.
travel costs at $20 per day per Lead Teacher	1,200.
lodging	n.a.
food and misc.: $10/day per Lead Teacher	600.
Total Cost per 2 Day Scoring Workshop	$33,000.

According to the worst case scenario, there will be a need for 7,099 trained scorers. Assuming these workshops produce 30 trained scorers, the cost of developing this network of scorer support will be $7,808,900.

(3) Continuing Scorer Training

Best Case Scenario

According to this scenario, teachers find scoring to be a

quite enjoyable and professionally enriching activity. They actively seek opportunities to learn how to do it, and once employed only rarely give up the job voluntarily. Moreover, there is considerable cross-over from the old tasks to the new so that there is a minimal need for formal retraining of those who continue.

To operationalize this view of the reality, I assume that what is required is the equivalent of 1/2 a day of a scorer's time to meet with a group of fellow scorers to discuss their activities. I envision a series of very small informal workshops where groups of scorers essentially teach and refresh themselves.

Cost of the 1/2 day 30 participant local district workshop:

$100 per diem for 1/2 day per scorer	50
travel:	n.a.
lodging:	n.a.
food and misc	n.a.

Total Number of Scorers according to the Best-Case Scenario: 37.5

Total Cost for Continuing Scorer Development:

37.5 * $50 = $1,875

Middle Case Scenario

If the middle-case scenario is accurate, there will be a moderate degree of turnover among scorers. Teachers are presumed to find scoring an interesting but demanding activity. It is presumed to be viewed positively but as a burden that needs to be shared equitably. Also, some degree of carry-over will be presumed to exist between old and new tasks, so that the teachers remaining as scorers require only modest amounts of new training.

I operationalize this scenario by assuming that the recurring training needs can be met with a one-day 30 participant regional workshop for 1/5 of the scoring cohort each year. The workshop will be taught by Lead Teachers and the ratio of participants to Lead Teachers will be 8:1.

Cost of a one-day regional workshop

$100 per diem for 1 day for 30 participants	3,000.
travel:$20 average for 30 participants (per day)	600.
lodging:	n.a.
materials ($10/participant)	300
food and misc $10/day per participant	300.
Lead Teacher costs	
assuming 3.7 instructors per 30 participant workshop	

$250 perdiem per Lead Teacher	925.
travel costs at $20 per day per instructor	74.
lodging	n.a.
food and misc.: $10/day per instructor	37.
Cost per 1 Day Continuing Staff Development Workshop	$5,236.

Total number of scorers in the middle-case scenario= 4,540

Total Cost for Continuing Staff Development:

(((.2) * (4,540)) / 30) * $5,236 = $158,476

Worst Case Scenario.

According to the worst case scenario, teachers will find scoring quite burdensome. They will avoid having to perform the service and they will seek to quit the job at the first opportunity. Thus, whatever efficiencies are gained thanks to experience will be lost because of the resulting high level of turnover. The high turnover will generate large and continuing demands for scorer training.

Moreover, this scenario holds that there will be little carry-over from prowess as a scorer with one set of tasks to performance as a scorer on new tasks that are developed. Thus, even those remaining on the job will need periodic training.

I assume that within this scenario, a training program for 1/3 of the scorer cohort will be required, on average, each year. This program will be divided into training for both new scorers who replace those exiting the system and "refresher-type" training for those who are continuing.

I assume that the magnitude of this program will correspond to the cost of a 2 full day regional workshop organized for 30 participants. I also assume that the Lead Teachers will serve as instructors and that the participant/Lead Teacher ratio will be 4:1. The costs of such a workshop are these:

Costs per 2 day Continuing Staff Development Workshop:

$100 per diem for 2 days for 30 participants	6,000.
travel:$20 average for 30 participants (per day)	1,200.
lodging:	n.a.
food and misc $10/day per participant	600.
materials ($20/participant)	600
Lead Teacher Costs	
assuming 7.5 Lead Teachers per 30 participants	
$250 perdiem per Lead Teacher	3,750.
travel costs at $20 per day per Lead Teacher	300.
lodging	n.a.
food and misc.: $10/day per Lead Teacher	150.

Total Cost per 2 Day Continuing Staff Development Workshop 12,600.

Total Number of Participants: (7,099) / 3 = 2,366

Total Cost of Providing Continuing Staff Development:

(2,366 / 30) * $12,600 = $993,720.

(4) Outside Auditing

I assume that one of the by-products of difficulty teaching teachers how to score will be a need for outside auditing; the greater the difficulty, the greater the need for outside auditing. The necessary auditing will not be confined to the performance tasks; the cumulative portfolios will also be subject to periodic audit.

Best Case Scenario

Here I assume that the Lead Teachers themselves can handle whatever auditing needs to be done. I also assume that they can do this during the equivalent of 1 full day per year. The implicit presumption is that the system works quite well and that only periodic spot checks are necessary. The Lead Teachers would not audit their own schools.

Cost of a 1 day block of time for 1 Auditor to Work

$250 per diem	250.
travel ($20/day)	20.
meals, etc. ($10/ day)	10.
lodging	n.a.
Total	280.

Recall that the number of Lead Teachers within the Best-Case Scenario is 110. If all of the Lead Teachers participate in the auditing phase of the project, the cost will be 110 * $280 = $30,800.

Middle Case Scenario

Within this scenario, auditing is a more serious problem. Again, I assume that all of the Lead Teachers are involved and that they need to meet and work the equivalent of 2 full days each.

Cost of a 2 day block of time for 1 Auditor to Work

$250 per diem (2 days)	500.
travel ($20/day)	40.
meals, etc. ($10/ day)	20.
lodging	n.a.
Total	$560.

Recall that the number of Lead Teachers according to the Middle-Case Scenario is 12,000. If the cost of the program is $560 per Lead Teacher and there are 12,000 Lead Teachers, the cost will be $6,720,000.

Worst Case Scenario

By assumption, the costs incurred to provide a relatively large amount of intensive training will not be sufficient to offset the difficulties teachers encounter as they seek to develop their scoring skills. I assume the training reduces but does not eliminate the problem. The failure to solve the problem through training necessitates the installation of a relatively extensive auditing system which will involve outside scorers routinely reviewing the performance exams and cumulative portfolios produced throughout the system. Double scoring will be commonplace. Perhaps even triple scoring.

Moreover, the public relations problems could be immense, particularly if the auditors are systematically lowering scores for a school, or if high stakes begin to be attached to these scores. These public relations needs can generate significant additional costs, but I will make no attempt here to estimate their magnitudes.

I continue to assume that the Lead Teachers can perform the auditing work but that they will each require the equivalent of 4

full days to accomplish their goals. I also assume that this work will require periodic regional meetings and therefore generates travel costs.

Cost of a 4 day period of time for 1 Auditor to Work

$250 per diem	1,000.
travel ($20/day)	80.
meals, etc. ($10/ day)	40.
lodging	n.a.
Total	1,120.

Recall that the number of Lead Teachers provided for within the worst-case scenario is 27,448. This implies an auditing cost of $30,741,760.

(5) Administration of Tasks

I have divided this section into two portions: A) Teacher Orientation, and B) Classroom Implementation. The Classroom Implementation section is also divided into two portions: 1) class time devoted to actual assessment, and 2) class time devoted to preparation.

A. TEACHER ORIENTATION

Best Case Scenario

My assumption here is that teachers will respond to the performance assessment approach quite readily. A 1/2 day 30 participant orientation program at the local level for all teachers that will be administering tasks and assembling cumulative student portfolios is all that is required. Note, however, that I am not dealing with whatever orientation might be necessary for teachers who are not directly involved in the administration of the exams (i.e., those at grade levels other than 4, 8, and 10). I assume a 30:1 ratio of participants to Lead Teachers. I also assume that these meetings will take place regionally. I have not provided an allowance for substitute teacher costs on the grounds that if the workshop takes place during regular school hours, the stipend paid to the teachers would logically be used to compensate the substitute teacher who is covering the teacher's class.

$100 per diem for .5 day	
for 30 participants	1,500.
travel:	n.a.
lodging:	n.a.
food and misc $10/day per participant	n.a.
instructor costs	
assuming 1 instructor per 30	
participant workshop	
$250 per diem per day per	

instructor	125.
travel costs at $20 per day	
per instructor	n.a.
lodging	n.a.
food and misc.: $10/day	
per instructor	n.a.
Total Cost per .5 Day Teacher Orientation Workshop for 30	$1,625.

The number of teachers requiring this orientation in a given year corresponds to the number of participating classrooms. Recall that under the terms of the best-case scenario, a total of 2,500 pupils will be assessed at each grade level each year (1,250 in each subject). This yields a total of 7,500 pupils. If there are 25 pupils in each class, this corresponds to a count of 300 classroom teachers. Assuming it costs $1,625 to orient a group of 30 teachers, the total cost of orientation will be $16,250.

Middle Case Scenario

Here I assume that the teacher orientation is less easily accomplished. In particular, I assume that the program requires the equivalent of a 1 day 30 participant regional workshop where the participant Lead Teacher ratio is 15:1.

The cost of such a workshop will be:

$100 per diem for 1 day	
for 30 participants	3,000.
travel:$20 average for 30 participants	600
lodging:	n.a.
food and misc $10/day per participant	300.
Lead Teacher Costs	
assuming 2 Lead Teachers per 30	
participants	
$250 per diem per day per	
Lead Teacher	500.
travel costs at $20 per day	
per Lead Teacher	40.
lodging	n.a.
food and misc.: $10/day	
per Lead Teacher	20.

Total Cost per 1 Day Teacher Orientation	
Workshop (for 30 participants)	$4,460.

The number of teachers requiring this orientation can be derived from the number of students being assessed under the terms of the middle-case scenario. These counts are: 124,000 4th grade students, 165,000 8th grade students, and 165,000 10th grade students. Assuming 25 students to a class and assuming the

participating 4th grade students are being assessed in both subjects by the same teacher, the number of 4th grade teachers requiring orientation will be 124,000/25 = 4,960. At the 8th and 10th grade levels, it is likely that participating students will be taught by two different teachers. Thus, if there are 165,000 8th grade students and if they are being assessed in 2 subject areas by different teachers and if the relevant pupil-teacher ratio is 25, the number of 8th grade teachers needing orientation will be (165,000/25) * 2 = 13,200. Similarly, 13,200 10th grade teachers will need to be oriented, according to this scenario. It follows that the total number of teachers requiring orientation will be 31,360.

If the cost for orienting 30 teachers is $4,460, the cost of orienting this number of teachers is (31,360/30) * $4,460 = $4,662,187.

Worst Case Scenario

Here the Lead Teachers fail in their effort to convey enthusiasm about performance assessment to their colleagues. Front-line teachers view performance assessment as a burden imposed on them by external authorities, and the Lead Teachers have no choice but to make a relatively intensive effort to orient teachers.

I assume that this translates into a need to provide the equivalent of a 2 full day workshop for every participating

teacher. Note: While I am costing this orientation in terms of a formal workshop, the reality is likely to be quite different with Lead Teachers working individually with front-line teachers.

I calculate the costs of mounting an orientation program for these teachers on the assumption that the workshop will be delivered regionally to groups of 30 teachers and that the relevant participant/instructor ratio is 7.5:1

Cost of a 2 day regional workshop for 30 participants

$100 per diem for 2 days	
for 30 participants	6,000.
travel:$20 average for 30 participants	1,200.
lodging:	n.a.
food and misc $10/day per participant	600.
Lead Teacher Costs	
assuming 4 Lead Teachers for 30 participants	
$250 per diem per day per Lead Teacher	2,000.
travel costs at $20 per day per Lead Teacher	160.
lodging	n.a.
food and misc.: $10/day per Lead Teacher	80.

Total Cost per 2 Day Teacher Orientation Workshop 10,040.

According to the Worst Case scenario, 254,200 4th grade, 227,865 8th grade, and 227,865 10th grade students need to be assessed. Again, assuming that the relevant pupil-teacher ratio is 25 and that the 4th grade teachers handle 2 subjects, 10,168 4th grade teachers will need orientation. With the same pupil-teacher ratio and assuming each teacher handles 1 subject at the 8th and 10th grade levels, 18,229 8th and 18,229 10th grade teachers will need orientation. The total number of teachers is 46,626.

If it costs $10,040 to orient 30 teachers, then the total cost of teacher orientation will be (46,626/30) * $10,040 = $15,604,302.

B. CLASSROOM IMPLEMENTATION

1)) Class Time Devoted to Actual Assessment

The assumption I impose here is that the amount of time teachers spend actually administering performance tasks will be the same regardless of whether it is a worst, middle, or best case scenario. For each grade level and subject, I assume that each task on average requires a total of 3 class hours. I also assume

that over the course of a year a student will complete 2 tasks.

Thus, for each class participating in the assessment, the time required will be 6 hours. In addition, I will assume that the teacher must spend 1 hour in preparation for each task. It follows that the teacher preparation time will be 2 hours, aside from the time spent being oriented.

The next step is to figure the cost, on average, of an hour of class time. I assume that the cost of an hour of teacher time is $25, and I adjust this figure upward by $5 to account for miscellaneous costs such as space, materials, utilities, and administrative overhead. Students' time is clearly required for the administration of performance assessment tasks, but there is no satisfactory means of recognizing its value in these cost calculations. For now, I note that students' time has value and is required by performance assessment activities, but I do not attempt to include estimates of its value in these cost calculations.

According to the best-case scenario, there will be 300 teachers that need to be oriented. This figure gives us a basis for assuming that the number of classes that will be involved in a given year will be 300. The corresponding figures for the middle-case and worst-case scenarios are: 31,360 and 46,626, respectively. Thus, the costs of actually administering the performance tasks are.

Best-Case 300 * 8 * $30 = $72,000.

Middle-Case	31,360	*	8	*	$30	=	$7,526,400.
Worst-Case	46,626	*	8	*	$30	=	$11,190,240.

2)) Class Time Devoted to Preparation

I assume that teachers take time from instruction to prepare their students for performance assessments. Again, there is a question about whether such time can simultaneously serve an instructional purpose, and I deal with this issue later in the treatment of cost absorption. For now I treat preparation time as a cost and I use the best, middle, and worst case scenarios to examine varying assumptions about how much time is devoted on average by teachers to preparation.

Best Case Scenario

Here my assumption is that .5 hour of preparation is spent for each 1 hour of class time devoted to performance assessment. The cost is: 300 * 6 * 0.5 * $30 = $27,000.

Middle Case Scenario

I assume here that 1.0 hours of preparation accompanies each hour of time devoted to performance assessment. Under this assumption, the costs of time devoted to class preparation will

be: 31,360 * 6 * 1.0 * $30 = $5,644,800.

Worst Case Scenario

I assume that for each hour of performance assessment, teachers within this scenario devote 1.5 hours of class time to preparation.[4] Under this assumption, the costs of time devoted to preparation will be:

46,626 * 6 * 1.5 * $30 = $12,589,020.

(6) Scoring

Best Case Scenario

Recall that there will be 15,000 tasks to score each year under the terms of the best-case scenario. There are 37.5 trained scorers in place, each handling 400 tasks. This requires 8 full days of work (50 tasks per day for 8 days). And I will assume that these scorers will be paid a stipend of $250 per day for this work.

Total scoring cost will be 37.5 * 8 * $250 = $75,000

---

[4] The 3:1 ratio between the best and worst case scenarios is not entirely arbitrary. The Office of Technology Assessment (1992, pg. 29) found that teachers in a large urban school district reported devoting up to 3 hours of preparation for each test administration. I am taking the upper figure here to reflect the worst-case scenario in terms of costs.

Middle Case Scenario

According to this scenario, there will be 4,540 scorers working 8 days at $250 per day. This yields a total scoring cost of 4,540 * 8 * $250 = $9,080,000.

Worst Case Scenario

The worst case (census approach) requires 7,099 scorers. The associated costs are: 7,099 * 8 * $250 = $14,198,000.

(7) Utilization of Results

It is important to include estimates of the costs associated with making use of the performance assessment results. I estimate these costs by making alternative assumptions about how much teacher time and Lead Teacher time will be required per hour of classroom time devoted to performance assessment.

Best Case Scenario

Here the teachers adapt quite readily to the use of performance assessment results. They require minimal supervision from Lead Teachers. I assume that for every hour of class time

devoted to performance assessment a teacher requires .12 of an hour of his/her time studying the results. I also assume that for every hour a classroom teacher devotes to reflecting on performance assessment results, .06 of an hour of Lead Teacher time will be required. This will be time spent working primarily one-to-one with the classroom teachers interpreting results and providing guidance.

Under these assumptions the costs of utilizing the results of performance assessment will be:

300 classes * 6 hours = 1,800 class-hours
1,800 class-hours * .12 = 216 additional teacher hours
216 * $30 = $6,480

In addition, the costs of the Lead Teachers' time need to be added.

216 * .06 = 12.96 hours

Assuming 8 hour days, this translates into 1.62 work days for Lead Teachers. Assuming their daily rate is $250, this involves an additional $405.
Total Cost = $6,480 + $405 = $6,885.

Middle Case Scenario

Here I assume that the teachers need to spend .25 hours for every hour of class time devoted to performance assessment, and that the Lead Teachers need to spend .12 of an hour for each teacher-hour devoted to interpretation.

31,360 classes * 6 hours = 188,160 class-hours
188,160 class-hours * .25 = 47,040 additional teacher hours
47,040 * $30 = $1,411,200

In addition, the costs of the Lead Teachers' time need to be added.

47,040 * .12 = 5,645 hours.

Assuming 8 hour days, this translates into 706 days of work for Lead Teachers. Assuming their daily rate is $250, this involves an additional $176,500.

Total Cost = $1,411,200 + $176,500 = $1,587,700.

Worst Case Scenario

Here teachers, on average, require considerable instruction and supervision in the utilization of performance assessment results. I assume that for every hour devoted to performance

assessment a teacher requires .5 of an hour of his/her time studying the results. I also assume that the for every hour a classroom teacher spends interpreting test results a Lead Teacher needs to spend .25 hours.

46,626 classes * 6 hours = 279,756 class-hours
279,756 class-hours * .50 = 139,878 additional teacher hours
139,878 * $30 = $4,196,340

In addition, the costs of the Lead Teachers' time need to be added.

139,878 * .25 = 34,970 hours

Assuming 8 hour days, this translates into 4,371 days of work for Lead Teachers. Assuming their daily rate is $250, this involves an additional $1,092,750.

Total Cost = $4,196,340 + $1,092,750 = $5,289,090.

(8) Administration and Overhead

There will be central administrative costs at both the national and individual state levels. The national costs will need to be spread across the various participating states and units. For now, I will conceive of central administrative support

as a State level matter. Contributions to the national level will be made out of the costs I enumerate below.

I will assume a flat $5 per participating pupil central administrative cost. In addition, I will consider costs associated with producing and distributing the examinations which serve as the basis of the performance assessment. The NSP has some experience with these costs and has found that production and distribution costs average $4.55 per participating pupil.[5]

Best Case Scenario

- There are 7,500 participating students

7,500	*	$5	$	37,500.
7,500	*	$4.55	$	34,125.
Total			$	71,625.

Middle Case Scenario

There are 454,000 participating students

[5] During the pilot testing, the NSP spent $300,000 to produce and distribute exams for a total of 2,640 classes of students (660 in each of two subjects and 2 grade levels). If there are 25 pupils in each class, 66,000 students were involved. The per participating student cost is 300,000 / 66,000 = $4.55.

454,000	*	$5	$2,270,000.
454,000	*	$4.55	$2,065,700.
Total			$4,335,700.

Worst Case Scenario

There are 709,930 participating students

709,930	*	$5	$3,549,650.
709,930	*	$4.55	$3,230,182
Total			$6,779,832

III. Alternative Assumptions About the Absorption of Costs

I have now completed a set of estimates for the operations costs of a performance assessment system for a large State where the assessment is focused on three grade levels in 2 subject areas. The costs I have totaled correspond to the dollar magnitudes of the ingredients that have been identified. No attention has been given to possible absorptions of costs through

the displacement of existing practice. In this final section of the cost analysis, I consider issues surrounding the possible absorption of the costs that have been enumerated. Recall that I dealt conceptually with this issue in Chapter 2. Once again, I make use of worst, middle, and best-case alternative scenarios.

In the following analyses, I make different assumptions about the magnitude of these absorptions. My rationale for doing so is that the assumed savings (however large they might be) occur because of the advent of performance assessment. Of course, it is possible that performance assessment occasions no savings or even generates additional costs at the local level. I explore the no savings result under the heading of the worst-case scenario. According to this view, performance assessment is a complete add-on and no local resources are released. I have not explored even more pessimistic scenarios, but the so-inclined reader is welcome to do so.

Within the middle and best-case scenarios I explore different views of how these savings could be realized. As the scenarios make clear, I see the potential for absorptions to arise in three areas: 1) local staff development; 2) the uses of class time for assessment (both preparation and the actual administration of the tasks); and 3) the utilization of assessment information.

(1) Local Staff Development

Worst Case Scenario

My presumption here is that resources currently being spent at the local level on in-service staff development are productive and there is no potential for absorbing the costs of teachers acquiring the skills associated with performance assessment. Thus, there is no adjustment necessary to the costs.

Middle Case Scenario

Here my presumption is that local school districts will welcome opportunities to orient their teachers in the uses of performance assessment. It will be viewed as a substitution of a productive use of staff development resources for uses which were highly questionable in terms of their impact on teacher performance.

The willingness of local districts to make this substitution reduces the level of new resources that need to be devoted to teacher orientation. I assume further that these savings generate a 50% reduction in the costs associated with Scorer Training, Continuing Scorer Training, and Classroom Teacher Orientation.

The revised figures are:

Scorer Training

Best	6,263	*	.5	=	$3,132
Middle	1,906,800	*	.5	=	$953,400
Worst	7,808,900	*	.5	=	$3,904,450

Continuing Scorer Training

Best	1,875	*	.50	=	$938
Middle	158,476	*	.50	=	$79,238
Worst	993,720	*	.50	=	$496,860

Teacher Orientation

Best	16,250	*	.5	=	$8,125
Middle	3,969,400	*	.5	=	$1,984,700
Worst	15,604,302	*	.5	=	$7,802,151

Best Case Scenario

I assume here a 75% absorption. The revised figures for teacher orientation are:

Scorer Training

Best	6,263	*	.25	=	$1,566
Middle	1,906,800	*	.25	=	$476,700
Worst	7,808,900	*	.25	=	$1,952,225

Continuing Scorer Training

Best	1,875	*	.25	=	$469
Middle	158,476	*	.25	=	$39,619
Worst	993,720	*	.25	=	$248,430

Teacher Orientation

Best	16,250	* .25 =		$4,063
Middle	3,969,400	* .25	=	$992,350
Worst	15,604,302	* .25	=	$3,901,076

(2) The Use of Classroom Time for Performance Assessment

There are two issues here. First there is the degree to which time devoted within classrooms to performance assessment can function as time devoted simultaneously to instruction. However, even if the time devoted to performance assessment can function in this way, there is still a cost to consider because the allocated time comes at the expense of time previously committed to instruction. In other words, students as a consequence learn less of some things and more of other things as a result of the introduction of performance assessment (assuming the total amount of classroom time remains unchanged).

The second issue concerns the comparative productivity of the two instructional uses of classroom time. It is only to the degree that time devoted to performance assessment is a <u>more</u> productive instructional use of time than what was done previously with the time, that you find a local potential to absorb a portion of the classroom time costs of performance assessment.

In the worst, middle, and best case scenarios below, I

explore the consequences surrounding different assumptions about the degree to which performance assessment uses of student time are more productive than alternative uses.

Worst Case Scenario

Within this scenario performance assessment is an add-on to existing classroom activities. The implicit presumption is that the previous uses of classroom time are productive. This view does not deny that performance assessment time can have instructional benefits, but the view presumes that there is no potential for local levels to absorb or offset the costs.

Middle Case Scenario

Here I assume that 50% of the costs of classroom time devoted to both administration and preparation can be absorbed locally. The underlying view is that schools at present are spending resources in classrooms in rather unproductive ways so that it is a matter of doing fewer things that have little or no payoff in exchange for the opportunity to do more of something that has a good payoff.

The revised figures for Classroom time costs are:

Task Administration

Best	$72,000	*	.5	=	$36,000

Middle	$7,526,400	*	.5	=	$3,763,200
Worst	$11,190,240	*	.5	=	$5,595,120

Class Preparation

Best	$27,000	*	.5	=	$13,500
Middle	$5,644,800	*	.5	=	$2,822,400
Worst	$12,589,020	*	.5	=	$6,294,510

Best Case Scenario

Here I assume the relevant rate of absorption is 75%. The revised figures for classroom time costs are:

Task Administration

Best	$72,000	*	.25	=	$18,000
Middle	$7,526,400	*	.25	=	$1,881,600
Worst	$11,190,240	*	.25	=	$2,797,560

Class Preparation

Best	$27,000	*	.25	=	$6,750
Middle	$5,644,800	*	.25	=	$1,411,200
Worst	$12,589,020	*	.25	=	$3,147,255

(3) The Utilization of Assessment Information.

The central question here is the degree to which the new assessment information actually makes a teacher's job easier. To the degree that the new information is easy to access and saves the teacher from devoting large amounts of time to pointless local testing activities, potentially large savings could be realized. These savings could even be larger than the cost of the time devoted to interpreting the results of the new assessments, thus giving rise to "negative costs."

Worst-Case Scenario.

No change is required here. The presumption is that there are no possible savings.

Middle Case Scenario

I assume a 50% rate of absorption.

The revised figures for the utilization of results are:

Best	$6,885	*	.5	=	$3,443
Middle	$1,587,700	*	.5	=	$793,850
Worst	$5,289,090	*	.5	=	$2,644,545

Best Case Scenario

The rate of absorption here is 75%. The revised figures are as follows:

Best	$6,885	* .25	=	$1,721
Middle	$1,587,700	* .25	=	$396,925
Worst	$5,289,090	* .25	=	$1,322,273

IV. Summary

Table 4-1 provides a summary of the Operations Costs examined in this section. The table covers a two year period. Year 5 includes the costs of training the scorers as well as the Supplemental Lead Teachers. Year 6 is the first fully operational year of the project; my assumption is that in Year 6 no new Lead Teachers and no new scorers need to be trained. Scorer costs in Year 6 and beyond are limited to the estimated costs of maintaining an appropriately sized cohort of trained scorers (see the earlier section that deal with Continuing Scorer Training).

Table 4-1 About Here

According to my estimates, the operations costs in a large State in Year 6 will range between a low of $209,000 and a high of $97.386 million. The middle case estimate is in the $24.858 to $48.673 million range, depending on how one wishes to treat the cost absorption issue. To place these numbers in some

perspective, if the pupil base for the state is 3.328,514 in grades K-12, the middle case operations costs expressed on a per pupil basis range between 11.72 and 7.47 dollars.

Table 4-2 summarizes the operations cost totals from Table 4-1 on a per pupil basis (using the 3,328,514 pupil count).

Table 4-2

Summary of Operations Costs in Year 6
For a Large State With 3.328 Million Pupils
(Costs/Pupil)

	Worst	Middle	Best
Best	.09	.07	.06
Middle	11.93	8.89	7.47
Worst	29.26	22.40	18.97

Note: The column headings refer to assumptions about the degree of cost absorption; the row headings refer to assumptions about the magnitude of program required to achieve the intended results. Cell entries are $/pupil.

CHAPTER 5

OPERATIONS COSTS IN A MID-SIZED STATE

I. Introduction

The focus in this chapter is on a mid-sized state where I assume there are 1,328 elementary schools and 374 secondary schools organized into 350 local education agencies (i.e., school districts). I will assume further that the state's grade 4 enrollment is 73,540 and that the enrollments in grades 8 and 10 are 70,402 and 71,117, respectively.

I deal explicitly with the following components of operations costs: (1) Supplemental Lead Teacher Training; (2) Scorer Training; (3) Continuing Scorer Training; (4) Outside Auditing; (5) Administration of Tasks; (6) Scoring; (7) Utilization of Results; and (8) Administration and Overhead (including the costs of printing and distributing the exams). Next I consider alternative assumptions regarding the possible absorptions of selected cost components. The chapter concludes with an overview and summary of the cost estimates.

II. Components of Cost

(1) Supplemental Lead Teacher Training

Available Supply of Lead Teachers

Recall that as a by-product of the Development Phase there is a pool of trained scorers. Under the assumptions I imposed in Chapter 3, I estimated the size of this pool per year to be 1,300 trained scorers per year (400 trained at the national scoring meetings and 900 trained regionally). At the end of the 4 year Development Phase, the maximum number of trained scorers will be 1,300 * 4 or 5,200, assuming there are no scorers who repeat their training program. This also presumes that there is no loss of skill over as long as 4 years for teachers who learn to score at the outset of the project.

Given the likelihood that scorers will vary in how well they learn the requisite skills, that some decay will take place over time for those who are trained early, and that the project will lose track of some participants, I will make alternative assumptions regarding the actual size of the reservoir of scorers that is available at the end of the Operations Phase.

According to the best-case scenario, there is little loss over time and teachers learn the relevant skills quite easily and uniformly. In other words, NSP does not have to deal with

significant unevenness in how well teachers learn to be scorers. Nor is there much unevenness in how well the trained teachers retain their skills. Nor does the project lose track of many scorers over time. The middle and worst-case scenarios relax these assumptions and introduce potentially significant levels of unevenness, depreciation and obsolescence, and loss.

There are no obvious benchmarks to rely upon in assigning magnitudes to the discount factors that need to be used, so I will make the relatively arbitrary assumption that under the best-case scenario, the effective loss is 10%, and that under the middle and worst-case scenarios, the respective percentage losses are 20 and 30.

My assumption is that these experienced scorers constitute the initial NSP representation in the field. These people will play lead roles in the training and implementation of the project within the participating states. They will be involved in both the performance assessment as well as the cumulative portfolio development aspects of the NSP. I will refer to them as Lead Teachers.[1]

I also assume that these Lead Teachers are divided across the participating states in proportion to the respective states'

---

[1] According to NSP documentation, performance tasks constitute just one part of the cumulative portfolio that will be generated for each student. While there will be central guidance provided about the types of items that should be included in students' portfolios, much discretion will be maintained at the individual school and teacher levels. The Lead Teachers will provide training and assistance to front-line teachers who are participating in the project.

populations. My rationale for this is based on the NSP practice of varying the number of invitations to the national scoring meetings according to its partners' populations (recall that either 2, 3, or 4 teachers from each grade level and subject were invited).

Recall that the size of this state is in the middle range with 1,328 elementary schools and 374 secondary schools with a total pupil enrollment in grades 4, 8, and 10 of 215,059. I assume that this state received 3 * 2 * 3 or 18 nationally trained scorers each year (during the Development Phase), and that the number of regional training workshops that were conducted within the state is proportional to the state's share of the NSP base student population (i.e., the population from all the participating states and units). According to NSP documentation, a state with 215,059 pupils would comprise 4.4 per cent of the pupil base being served by the project. Thus, I assume that it operated 4.4 per cent of the 30 regional workshops that were held each year. This corresponds to 1.32 workshops per year. Recall that each workshop generated 30 trained scorers. It follows that for each year the mid-sized State created a pool of 1.32 * 30 or 39.6 locally trained scorers. This yields a total of 18 + 39.6 or 57.6 potential Lead Teachers each year for a total possible of 230.4. The application of the best, middle, and worst case loss rates that I derived above generates the following estimates of Lead Teacher Supply for the mid-size State at the close of the Development Phase of the Project:

Best Case

$$230.4 - (.1 * 230.4) = 207.36$$

Middle Case

$$230.4 - (.2 * 230.4) = 184.32$$

Worst Case

$$230.4 - (.3 * 230.4) = 161.28$$

Demand for Lead Teachers During Operations

The next question is whether the supply of these Lead Teachers thanks to the Development Phase of the Project will be adequate to staff the Operations Phase of the Project. To begin to answer this question, I make a series of assumptions about the scope of the operational phase of the performance assessment project. There are two dimensions to this demand: (1) the number of schools that will be involved in the operational version of performance assessment; and (2) the level of direct supervision by Lead Teachers that is required within each participating school. I will be making alternative assumptions regarding each dimension, and I shall join them within the scenario framework so that the best case of one is linked with the best case of the other. In other words, I will not be considering alternative combinations of best, middle, and worst cases along each dimension.

Counts of Participating Schools

At one extreme, I will assume that in order for the project to achieve its goals, it will be necessary to implement an annual performance assessment program within every school in each participating state. I call this a census approach to implementation, and it corresponds to a worst-case scenario with respect to the associated costs.

Recall that a major goal of the NSP is to change fundamentally the conduct of instruction throughout entire schooling systems. According to this worst-case cost scenario, it is necessary to have an NSP presence within every school during every year of the operations phase of the project. I also make alternative assumptions about the level of the presence that is required, but for now the focus is on how many schools need to participate in a given year during the operational phase.

At the other extreme, I will assume that it is possible for the project to achieve its goals through the use of a light matrix sampling design. The presumption here will be that a periodic program of assessment within a relatively small sample of schools is sufficient within each state to achieve the far-reaching goals of the NSP. The sample of schools and classrooms participating will vary from year to year. All schools and the relevant classrooms will be eligible for selection, and at any given time teachers and administrators will not know when their classrooms

and schools will participate. Moreover, in any given year, I assume that the state will focus on some subset of the possible tasks within the Task Bank. This scenario will correspond to a best-case view of costs since fewer resources will be required (by assumption) for the project to achieve its goals.

The middle case scenario involves a situation where there is interest in district specific results. In contrast to the census and matrix approaches, the presumption here is that there is interest in district level performance. The design will require sampling from within districts and this will require a measure of performance assessment that lies between the first two extremes that I have identified.

Level of Direct Supervision Provided By Lead Teachers

According to the NSP proposal, a goal of the project is to have two externally trained and certified scorers within each school participating in the performance assessment activities. I am assuming that such people correspond to what I have called Lead Teachers, and I note that there is some ambiguity surrounding the precise level of Lead Teacher supervision that will be appropriate. At one extreme, it could be that two Lead Teachers could handle all of the testing taking place within a school regardless of the subject being taught. Thus, in a secondary school with grades 8 and 10 present, two Lead Teachers could handle the testing program for both mathematics and language arts.

From a cost perspective, this extreme corresponds to a best-case scenario.

At the opposite extreme, it may be necessary to have two Lead Teachers for each grade and subject being assessed. In this case, a secondary school with two grade levels would require 8 Lead Teachers. This reality corresponds to a worst-case scenario in terms of costs.

A middle ground can be defined by thinking of the Lead Teachers as being able to cross grade levels but not subject areas. Under the terms of this middle-case scenario, the secondary school with grades 8 and 10 would require 4 Lead Teachers.

These three scenarios (for both the number of participating schools and the number of needed Lead Teachers in each school) are used below to define the demand for Lead Teachers in a typical operational year of the project.

Best Case Scenario

Number of participating schools. This scenario involves the use of a matrix sampling design. I assume that the sampling goal will be 100 observations per task,[2] and that in any given year the State will employ 25 per cent of the tasks available within the

[2] According to the NSP, a sample of 2,500 observations needs to be drawn for all 25 tasks for each subject at each grade level being considered to satisfy psychometric concerns over validity.

Task Bank. If every student participating in the program received one task, implementation would require 25 * 100 or 2,500 pupils per grade per subject. But, I will also assume that each student participating will complete 2 tasks, and thereby reduce the required number of participating students by half to 1,250 per grade per subject.

Recall that the mid-size State has a grade 4 enrollment of 73,540 and a population of 1,328 elementary schools. If the grade 4 students are evenly distributed across the schools, it follows that the average school will enroll approximately 55 4th grade students. If the goal is to have 2,500 participating 4th grade students (1,250 per subject), in a given year performance assessment will need to take place within approximately 45 of the 1,328 elementary schools (2,500/55).

At the secondary level, there are two grade levels. For the mid-size State, there are 70,402 8th grade students and 71,117 10th grade students. There are 374 secondary schools and assuming all 8th grade students are enrolled in secondary schools and that the students are evenly distributed across the schools, it follows that each school enrolls, on average, approximately 189 students at each of these grade levels. If the goal is to have 2,500 participating 8th and 2,500 participating 10th grade students (again, 1,250 per subject), then in a given year performance assessments will need to take place within approximately 13 of the 374 secondary schools (2,500/189).

Level of staffing. In keeping with the best-case scenario, I

assume that each participating school needs 2 Lead Teachers and that these Lead Teachers can handle both multiple subject areas and grade levels (where they occur). If there are 45 elementary schools in the program, there will need to be (45*2) or 90 Lead Teachers for the elementary schools. If there are 13 secondary schools in the program, there will need to be (13*2) or 26 Lead Teachers for the secondary schools.

Thus, the best-case scenario involves a total annual demand of 90+26= 116 Lead Teachers. This compares with the derived supply of 207.36. Thus, under terms of the best-case scenario, the mid-size State will not need to provide supplemental training for Lead Teachers, at least not at the outset of operations. The costs of supplemental Lead Teacher training will be considered 0 for the best case scenario.

Middle Case Scenario

Number of participating schools. Here the idea is that the state is interested in having information from each district, and the presumption is that the matrix sampling design described above misses a significant number of districts. As I indicated earlier, the mid-size State operates 350 separate school districts. I assume that the average grade 4 enrollment within each district is 73,540 / 350 or 210 and that the average grade 8 and 10 enrollments are 70,402 / 350 or 201 and 71,117 / 350 or 203, respectively. Using the 55 4th grade pupils per school and 189

8th or 10th grade pupils per school figures that I derived above, it follows that on average each district operates 3.8 elementary schools and 1.1 secondary schools.

I assume that a sample of 2 elementary schools per district and 1 secondary school per district will be adequate to provide the district level aggregates. This will require staffing performance assessment activities in 700 elementary schools and 350 secondary schools in a given year.

Level of staffing. In accordance with the middle-case scenario where the assumption is that 2 Lead Teachers are needed for each subject within each school, there is an implied demand 700 * 2 * 2 = 2,800 Lead Teachers for the elementary program and 350 * 2 * 2 = 1,400 Lead Teachers for the secondary program.[3] This means the state needs a pool of 4,200 Lead Teachers compared to the 184 that are available following the Development Phase.

Implications for costs. I assume the necessary training will take the form of a supplemental series of 4 day workshops structured around scoring exercises. The same costs that I derived earlier will apply. Recall that these workshops cost $18,810 and yielded 30 trained scorers. Thus, the supplemental cost for Lead Teacher training for the mid-size State according to the middle case scenario will be:

---

[3] Recall that the assumption is that Lead Teachers can cross grade levels but not subject areas. This explains why the secondary schools require 4 rather than 8 Lead Teachers.

((4,200 - 184) / 30) * $18,810 = $2,518,032

Worst Case Scenario

Number of participating schools. Recall that the worst-case scenario involves a census approach to performance assessment where the presumption is that every 4th, 8th, and 10th grade student needs to be assessed every year in both subject areas.

The mid-sized State has a population of 1,328 elementary and 374 secondary schools. If the state pursues a census approach, Lead Teacher staffing will be required in each of these schools.

Level of staffing. According to the worst-case scenario, 2 Lead Teachers are needed for each possible combination of subject and grade level. Assuming elementary schools involve only grade 4, the total number of Lead Teachers needed for the 4th grade assessment program will be 1,328 * 2 * 2 or 5,312. The corresponding number of Lead Teachers for the 8th and 10th grade assessment programs (assuming they are all located within the secondary schools) will be 374 * 2 * 2 * 2 or 2,992.

Implications for costs. The total number of Lead Teachers needed according to this scenario is 5,312 + 2,992 = 8,340. In contrast, according to the worst-case scenario, the Development Phase of the project generates a supply of 161 Lead Teachers. The relevant cost calculation (assuming the Lead Teachers are trained through the use of regional workshops) is:

((8,340 - 161) / 30) * 18,810 = $5,128,233

(2) Scorer Training

Best Case Scenario

Number of Scorers Needed

Assuming each participating student generates 2 tasks, the annual total number of tasks that need to be scored will be the number of students per grade level (2,500) * the number of grade levels (3) * the number of tasks completed (2) = 15,000.

I assume that each scorer scores 400 tasks. This is the equivalent of 8 days of work. The NSP does not seek to develop a supply of "professional" task scorers. It is, instead, committed to achieving a broad base of participation among teachers and others. For this reason, I impose the 400 task ceiling.

If there are 15,000 tasks to score in the large State, and if each scorer scores 400, the demand for scorers will be 37.5.

Level of Training Required

Minimal training will be required to train local scorers under terms of the best-case scenario. The underlying assumption is that this kind of assessment and its scoring will be very much in-line with how teachers think and go about their work. The teachers are presumed to adapt quickly and easily. I assume that

the training can be done quite informally within the local districts; as a consequence travel costs for participants become negligible and will be omitted. Since the Lead Teachers will be traveling, I have included an allowance for this travel in the budget.

However, for the sake of deriving cost estimates, I will continue to treat the training as if it has a group workshop nature. In particular, I will assume that what is necessary is the equivalent of a one-day workshop for 30 participants where the participant/Lead Teacher ratio is 8:1.

$100 per diem for 1 day for 30 participants	3,000.
travel:$20 average for 30 participants	n.a.
lodging:	n.a.
food and misc $10/day per participant	300.
materials (20/participant)	600.
Lead Teacher costs	
assuming 3.7 Lead Teachers per 30 participants	
$250 per diem per day per Lead Teacher	925.
travel costs at $40 per Lead Teacher	148.
lodging	n.a.
food and misc.: $10/day per instructor	37.
Total Cost per 1 Day Scoring Workshop	$5,010.

The yield for this workshop is 30 trained scorers. If the need for scorers is 37.5, the costs of training these individuals will be $6,263.

Middle Case Scenario

Number of Scorers Needed

According to this scenario, there will be assessment activities in 700 elementary schools and 350 high schools. The average number of 4th grade students per elementary school is 55; the average number of 8th and 10th grade students is 189. Thus, there are 170,800 students participating in a given year. If each student completes 4 tasks (two for each of two subjects), there will be 683,200 tasks to score.

If scorers score 400 tasks each, there will be a demand for 1,708 scorers.

Level of Training Required

A more ambitious level of training is required under the terms of the middle-case scenario. Instead of the equivalent of a 1 day (30 person) scoring workshop, I will assume that 2 days are necessary. I shall also assume that a more intensive training

experience is necessary. Instead of the 8:1 ratio of participants to Lead Teachers, I will assume that a 4:1 ratio is necessary. I shall also build travel costs into the budget, since my presumption is that it will be less possible for the training to place informally at the home sites.

$100 per diem for 2 days for 30 participants	6,000.
travel:$20 average for 30 participants (per day)	1,200.
lodging:	n.a.
food and misc $10/day per participant	600.
materials (20/participant)	600.
Lead Teacher costs	
assuming 7.5 Lead Teachers per 30 participants	
$250 per diem per day per Lead Teacher	3,750.
travel costs at $20 per day per instructor	300.
lodging	n.a.
food and misc.: $10/day per instructor	150.
Total Cost per 2 Day Scoring Workshop	$12,600.

Assuming there are 1,708 scorers that need to be trained, and assuming these 2 day workshops each yield 30 trained scorers, the cost of scorer training will be $717,360.

Worst Case Scenario

Number of Scorers Needed

The mid-sized State operates 1,328 elementary schools and 374 secondary schools. Assuming there are 55 4th grade students per elementary school and 189 8th and 10th grade students per secondary school, there will be 214,412 students being assessed each year. If each student completes 4 tasks, there will be 857,648 tasks to score each year.

Assuming each scorer scores 400 tasks, there will be a demand for 2,144 scorers.

Level of Training Required

Since this is the worst case scenario, I assume that teachers, on balance, find it difficult to grasp the requisite skills to function effectively as scorers. I assume that these teachers need to spend the equivalent of 4 one-day workshops acquiring these skills, and that these workshops will be offered regionally. I assume further that the scorer/Lead Teacher ratio in the workshop needs to be 2:1.

For each 4 one-day elementary task scoring workshop, there will be the following costs:

$100 per diem for 4 days for 30 participants	12,000.
travel:$20 average for 30 participants (per day)	2,400.
lodging:	n.a.
food and misc $10/day per participant	1,200.
materials (20/participant)	600.
Lead Teacher costs	
assuming 15 Lead Teachers per 30 participants	
$250 per diem per day per Lead Teacher	15,000.
travel costs at $20 per day per Lead Teacher	1,200.
lodging	n.a.
food and misc.: $10/day per Lead Teacher	600.
Total Cost per 2 Day Scoring Workshop	$33,000.

According to the worst case scenario, there will be a need for 2,144 trained scorers. Assuming these workshops produce 30 trained scorers, the cost of developing this network of scorer support will be $2,358,532.

(3) Continuing Scorer Training

Best Case Scenario

According to this scenario, teachers find scoring to be a quite enjoyable and professionally enriching activity. They actively seek opportunities to learn how to do it, and once employed only rarely give up the job voluntarily. Moreover, there is considerable cross-over from the old tasks to the new so that there is a minimal need for formal retraining of those who continue.

To operationalize this view of the reality, I assume that what is required is the equivalent of 1/2 a day of a scorer's time to meet with a group of fellow scorers to discuss their activities. I envision a series of very small informal workshops where groups of scorers essentially teach and refresh themselves.

Cost of the 1/2 day 30 participant local district workshop:

$100 perdiem for 1/2 day per scorer	50
travel:	n.a.
lodging:	n.a.
food and misc	n.a.

Total Number of Scorers according to the Best-Case Scenario: 37.5

Total Cost for Continuing Scorer Development:

37.5 * $50 = $1,875

Middle Case Scenario

If the middle-case scenario is accurate, there will be a moderate degree of turnover among scorers. Teachers are presumed to find scoring an interesting but demanding activity. It is presumed to be viewed positively but as a burden that needs to be shared equitably. Also, some degree of carry-over will be presumed to exist between old and new tasks, so that the teachers remaining as scorers require only modest amounts of new training.

I operationalize this scenario by assuming that the recurring training needs amount to a one-day 30 participant regional workshop for 1/5 of the scoring cohort each year. The workshop will be taught by Lead Teachers and the ratio of participants to Lead Teachers will be 8:1.

Cost of a one-day regional workshop

$100 perdiem for 1 day for 30 participants	3,000.
travel:$20 average for 30 participants (per day)	600.
lodging:	n.a.
materials ($10/participant)	300
food and misc $10/day per participant	300.
Lead Teacher costs	
assuming 3.7 instructors per 30 participant workshop	

$250 perdiem per Lead Teacher	925.
travel costs at $20 per day per instructor	74.
lodging	n.a.
food and misc.: $10/day per instructor	37.
Cost per 1 Day Continuing Staff Development Workshop	$5,236.

Total number of scorers
in the middle-case scenario = 1,708

Total Cost for Continuing Staff Development:

(((.2) * (1,708)) / 30) * $5,236 = $59,621.

Worst Case Scenario.

According to the worst case scenario, teachers will find scoring quite burdensome. They will avoid having to perform the service and they will seek to quit the job at the first opportunity. Thus, whatever efficiencies are gained thanks to experience will be lost because of the resulting high level of turnover. The high turnover will generate large and continuing demands for scorer training.

Moreover, this scenario holds that there will be little carry-over from prowess as a scorer with one set of tasks to performance as a scorer on new tasks that are developed. Thus, even those remaining on the job will need periodic training.

I assume that within this scenario, a training program for 1/3 of the scorer cohort will be required, on average, each year. This program will be divided into training for both new scorers who replace those exiting the system and "refresher-type" training for those who are continuing.

I assume that the magnitude of this program will correspond to the cost of a 2 full day regional workshop organized for 30 participants. I also assume that the Lead Teachers will serve as instructors and that the participant/Lead Teacher ratio will be 4:1. The costs of such a workshop are these:

Costs per 2 day Continuing Staff Development Workshop:

$100 per diem for 2 days for 30 participants	6,000.
travel:$20 average for 30 participants (per day)	1,200.
lodging:	n.a.
food and misc $10/day per participant	600.
materials ($20/participant)	600
Lead Teacher Costs	
assuming 7.5 Lead Teachers per 30 participants	
$250 perdiem per Lead Teacher	3,750.
travel costs at $20 per day per Lead Teacher	300.
lodging	n.a.
food and misc.: $10/day	

per Lead Teacher	150.
Total Cost per 2 Day Continuing Staff Development Workshop	12,600.

Total Number of Participants: (2,144) / 3 = 715

Total Cost of Providing Continuing Staff Development:

(715 / 30) * $12,600 = $300,300.

(4) Outside Auditing

I assume that one of the by-products of difficulty teaching teachers how to score will be a need for outside auditing; the greater the difficulty, the greater the need for outside auditing. The necessary auditing will not be confined to the performance tasks; the cumulative portfolios will also be subject to periodic audit.

Best Case Scenario

Here I assume that the Lead Teachers themselves can handle whatever auditing needs to be done. I also assume that they can do this during the equivalent of 1 full day per year. The implicit presumption is that the system works quite well and that only periodic spot checks are necessary. The Lead Teachers would not audit their own schools.

Cost of a 1 day block of time for 1 Auditor to Work

$250 per diem	250.
travel ($20/day)	20.
meals, etc. ($10/ day)	10.
lodging	n.a.
Total	280.

Recall that the number of Lead Teachers within the Best-Case Scenario is 116. If all of the Lead Teachers participate in the auditing phase of the project, the cost will be 116 * $280 = $32,480.

Middle Case Scenario

Within this scenario, auditing is a more serious problem. Again, I assume that all of the Lead Teachers are involved and that they need to meet and work the equivalent of 2 full days each.

Cost of a 2 day block of time for 1 Auditor to Work

$250 per diem (2 days)	500.
travel ($20/day)	40.
meals, etc. ($10/ day)	20.
lodging	n.a.

Total	$560.

Recall that the number of Lead Teachers according to the Middle-Case Scenario is 4,200. If the cost of the program is $560 per Lead Teacher and there are 4,200 Lead Teachers, the cost will be $2,352,000.

Worst Case Scenario

By assumption, the costs incurred to provide a relatively large amount of intensive training will not be sufficient to offset the difficulties teachers encounter as they seek to develop their scoring skills. I assume the training reduces but does not eliminate the problem. The failure to solve the problem through training necessitates the installation of a relatively extensive auditing system which will involve outside scorers routinely reviewing the performance exams and cumulative portfolios produced throughout the system. Double scoring will be commonplace. Perhaps even triple scoring.

Moreover, the public relations problems could be immense, particularly if the auditors are systematically lowering scores for a school, or if high stakes begin to be attached to these scores. These public relations needs can generate significant additional costs, but I will make no attempt here to estimate their magnitudes.

I continue to assume that the Lead Teachers can perform the auditing work but that they will each require the equivalent of 4 full days to accomplish their goals. I also assume that this work will require periodic regional meetings and therefore generates travel costs.

Cost of a 4 day period of time for 1 Auditor to Work

$250 perdiem	1,000.
travel ($20/day)	80.
meals, etc. ($10/ day)	40.
lodging	n.a.
Total	1,120.

Recall that the number of Lead Teachers provided for within the worst-case scenario is 8,340. This implies an auditing cost of $9,340,800.

(5) Administration of Tasks

I have divided this section into two portions: A) Teacher Orientation, and B) Classroom Implementation. The Classroom Implementation section is also divided into two portions: 1) class time devoted to actual assessment, and 2) class time devoted to preparation.

A. TEACHER ORIENTATION

Best Case Scenario

My assumption here is that teachers will respond to the performance assessment approach quite readily. A 1/2 day 30 participant orientation program at the local level for all teachers that will be administering tasks and assembling cumulative student portfolios is all that is required. Note, however, that I am not dealing with whatever orientation might be necessary for teachers who are not directly involved in the administration of the exams (i.e., those at grade levels other than 4, 8, and 10). I assume a 30:1 ratio of participants to Lead Teachers. I also assume that these meetings will take place regionally. I have not provided an allowance for substitute-teacher costs on the grounds that if the workshop takes place during regular school hours, the stipend paid to the teachers would logically be used to compensate the substitute teacher who is covering the teacher's class.

$100 per diem for .5 day	
for 30 participants	1,500.
travel:	n.a.
lodging:	n.a.
food and misc $10/day per participant	n.a.
instructor costs	
assuming 1 instructor per 30	

participant workshop	
$250 per diem per day per instructor	125.
travel costs at $20 per day per instructor	n.a.
lodging	n.a.
food and misc.: $10/day per instructor	n.a.
Total Cost per .5 Day Teacher Orientation Workshop for 30	$1,625.

The number of teachers requiring this orientation in a given year corresponds to the number of participating classrooms. Recall that under the terms of the best-case scenario, a total of 2,500 pupils will be assessed at each grade level each year (1,250 in each subject). This yields a total of 7,500 pupils. If there are 25 pupils in each class, this corresponds to a count of 300 classroom teachers. Assuming it costs $1,625 to orient a group of 30 teachers, the total cost of orientation will be $16,250.

Middle Case Scenario

Here I assume that the teacher orientation is less easily accomplished. In particular, I assume that the program requires

the equivalent of a 1 day 30 participant regional workshop where the participant Lead Teacher ratio is 15:1.

The cost of such a workshop will be:

$100 per diem for 1 day	
for 30 participants	3,000.
travel:$20 average for 30 participants	600
lodging:	n.a.
food and misc $10/day per participant	300.
Lead Teacher Costs	
assuming 2 Lead Teachers per 30 participants	
$250 per diem per day per Lead Teacher	500.
travel costs at $20 per day per Lead Teacher	40.
lodging	n.a.
food and misc.: $10/day per Lead Teacher	20.
Total Cost per 1 Day Teacher Orientation Workshop (for 30 participants)	$4,460.

The number of teachers requiring this orientation can be derived from the number of students being assessed under the terms of the middle-case scenario. These counts are: 38,500 4th grade

students, 66,150 8th grade students, and 66,150 10th grade students. Assuming 25 students to a class and assuming the participating 4th grade students are being assessed in both subjects by the same teacher, the number of 4th grade teachers requiring orientation will be 38,500/25 = 1,540. At the 8th and 10th grade levels, it is likely that participating students will be taught by two different teachers. Thus, if there are 66,150 8th grade students and if they are being assessed in 2 subject areas by different teachers and if the relevant pupil-teacher ratio is 25, the number of 8th grade teachers needing orientation will be (66,150/25) * 2 = 5,292. Similarly, 5,292 10th grade teachers will need to be oriented, according to this scenario. It follows that the total number of teachers requiring orientation will be 12,124.

If the cost for orienting 30 teachers is $4,460, the cost of orienting this number of teachers is (12,124/30) * $4,460 = $1,802,435.

Worst Case Scenario

Here the Lead Teachers fail in their effort to convey enthusiasm about performance assessment to their colleagues. Front-line teachers view performance assessment as a burden imposed on them by external authorities, and the Lead Teachers have no choice but to make a relatively intensive effort to orient teachers.

I assume that this translates into a need to provide the equivalent of a 2 full day workshop for every participating teacher. Note: While I am costing this orientation in terms of a formal workshop, the reality is likely to be quite different with Lead Teachers working individually with front-line teachers.

I calculate the costs of mounting an orientation program for these teachers on the assumption that the workshop will be delivered regionally to groups of 30 teachers and that the relevant participant/instructor ratio is 7.5:1

Cost of a 2 day regional workshop for 30 participants

$100 per diem for 2 days	
for 30 participants	6,000.
travel:$20 average for 30 participants	1,200.
lodging:	n.a.
food and misc $10/day per participant	600.
Lead Teacher Costs	
assuming 4 Lead Teachers for 30 participants	
$250 per diem per day per Lead Teacher	2,000.
travel costs at $20 per day per Lead Teacher	160.
lodging	n.a.

food and misc.: $10/day per Lead Teacher	80.
Total Cost per 2 Day Teacher Orientation Workshop	10,040.

According to the Worst Case scenario, 73,040 4th grade, 70,686 8th grade, and 70,686 10th grade students need to be assessed. Again, assuming that the relevant pupil-teacher ratio is 25 and that the 4th grade teachers handle 2 subjects, 2,922 4th grade teachers will need orientation. With the same pupil-teacher ratio and assuming each teacher handles 1 subject at the 8th and 10th grade levels, 5,655 8th and 5,655 10th grade teachers will need orientation. The total number of teachers is 14,232.

If it costs $10,040 to orient 30 teachers, then the total cost of teacher orientation will be (14,232/30) * $10,040 = $4,762,976.

B. CLASSROOM IMPLEMENTATION

1)) Class Time Devoted to Actual Assessment

The assumption I impose here is that the amount of time teachers spend actually administering performance tasks will be the same regardless of whether it is a worst, middle, or best case scenario. For each grade level and subject, I assume that each task on average requires a total of 3 class hours. I also assume

that over the course of a year a student will complete 2 tasks.

Thus, for each class participating in the assessment, the time required will be 6 hours. In addition, I will assume that the teacher must spend 1 hour in preparation for each task. It follows that the teacher preparation time will be 2 hours, aside from the time spent being oriented.

The next step is to figure the cost, on average, of an hour of class time. I assume that the cost of an hour of teacher time is $25, and I adjust this figure upward by $5 to account for miscellaneous costs such as space, materials, utilities, and administrative overhead. Students' time is clearly required for the administration of performance assessment tasks, but there is no satisfactory means of recognizing its value in these cost calculations. For now, I note that students' time has value and is required by performance assessment activities, but I do not attempt to include estimates of its value in these cost calculations.

According to the best-case scenario, there will be 300 teachers that need to be oriented. This figure gives us a basis for assuming that the number of classes that will be involved in a given year will be 300. The corresponding figures for the middle-case and worst-case scenarios are: 12,124 and 14,232, respectively. Thus, the costs of actually administering the performance tasks are:

Best Case 300 * 8 * $30 = $72,000.

Middle Case	12,124	*	8	*	$30	=	$2,909,760.
Worst Case	14,232	*	8	*	$30	=	$3,415,680.

2)) Class Time Devoted to Preparation

I assume that teachers take time from instruction to prepare their students for performance assessments. Again, there is a question about whether such time can simultaneously serve an instructional purpose, and I deal with this issue later in the treatment of cost absorption. For now I treat preparation time as a cost and I use the best, middle, and worst case scenarios to examine varying assumptions about how much time is devoted on average by teachers to preparation.

Best Case Scenario

Here my assumption is that .5 hour of preparation is spent for each 1 hour of class time devoted to performance assessment. The cost is: 300 * 6 * 0.5 * $30 = $27,000.

Middle Case Scenario

I assume here that 1.0 hours of preparation accompanies each hour of time devoted to performance assessment. Under this assumption, the costs of time devoted to class preparation will

be: 12,124 * 6 * 1.0 * $30 = $2,182,320.

Worst Case Scenario

I assume that for each hour of performance assessment, teachers within this scenario devote 1.5 hours of class time to preparation.[4] Under this assumption, the costs of time devoted to preparation will be:

14,232 * 6 * 1.5 * $30 = $3,842,640.

(6) Scoring

Best Case Scenario

Recall that there will be 15,000 tasks to score each year under the terms of the best-case scenario. There are 37.5 trained scorers in place, each handling 400 tasks. This requires 8 full days of work (50 tasks per day for 8 days). And I will assume that these scorers will be paid a stipend of $250 per day for this work.

Total scoring cost will be 37.5 * 8 * $250 = $75,000

---

4 The 3:1 ratio between the best and worst case scenarios is not entirely arbitrary. The Office of Technology Assessment (1992, pg. 29) found that teachers in a large urban school district reported devoting up to 3 hours of preparation for each test administration. I am taking the upper figure here to reflect the worst-case scenario in terms of costs.

Middle Case Scenario

According to this scenario, there will be 1,708 scorers working 8 days at $250 per day. This yields a total scoring cost of 1,708 * 8 * $250 = $3,416,000.

Worst Case Scenario

The worst case (census approach) requires 2,144 scorers. The associated costs are: 2,144 * 8 * $250 = $4,288,000.

(7) Utilization of Results

It is important to include estimates of the costs associated with making use of the performance assessment results. I estimate these costs by making alternative assumptions about how much teacher time and Lead Teacher time will be required per hour of classroom time devoted to performance assessment.

Best Case Scenario

Here the teachers adapt quite readily to the use of performance assessment results. They require minimal supervision from Lead Teachers. I assume that for every hour of class time devoted to performance assessment a teacher requires .12 of an

hour of his/her time studying the results. I also assume that for every hour a classroom teacher devotes to reflecting on performance assessment results, .06 of an hour of Lead Teacher time will be required. This will be time spent working primarily one-to-one with the classroom teachers interpreting results and providing guidance.

Under these assumptions the costs of utilizing the results of performance assessment will be:

300 classes * 6 hours = 1,800 class-hours
1,800 class-hours * .12 = 216 additional teacher hours
216 * $30 = $6,480

In addition, the costs of the Lead Teachers' time need to be added.

216 * .06 = 12.96 hours

Assuming 8 hour days, this translates into 1.62 work days for Lead Teachers. Assuming their daily rate is $250, this involves an additional $405.

Total Cost = $6,480 + $405 = $6,885.

Middle Case Scenario

Here I assume that the teachers need to spend .25 hours for every hour of class time devoted to performance assessment, and that the Lead Teachers need to spend .12 of an hour for each teacher-hour devoted to interpretation.

12,124 classes * 6 hours = 72,744 class-hours
72,744 class-hours * .25 = 18,186 additional teacher hours
18,186 * $30 = $545,580

In addition, the costs of the Lead Teachers' time need to be added.

18,186 * .12 = 2,182 hours.

Assuming 8 hour days, this translates into 273 days of work for Lead Teachers. Assuming their daily rate is $250, this involves an additional $68,250.

Total Cost = $545,580 + $68,250 = $613,830.

Worst Case Scenario

Here teachers, on average, require considerable instruction and supervision in the utilization of performance assessment results. I assume that for every hour devoted to performance assessment a teacher requires .5 of an hour of his/her time

studying the results. I also assume that the for every hour a classroom teacher spends interpreting test results a Lead Teacher needs to spend .25 hours.

14,232 classes * 6 hours = 85,392 class-hours
85,392 class-hours * .50 = 42,696 additional teacher hours
42,696 * $30 = $1,280,880.

In addition, the costs of the Lead Teachers' time need to be added.

42,696 * .25 = 10,674 hours

Assuming 8 hour days, this translates into 1,334 days of work for Lead Teachers. Assuming their daily rate is $250, this involves an additional $333,500.

Total Cost = $1,280,880 + $333,500 = $1,614,380.

(8) Administration and Overhead

There will be central administrative costs at both the national and individual state levels. The national costs will need to be spread across the various participating states and units. For now, I will conceive of central administrative support as a State level matter. Contributions to the national level will

be made out of the costs I enumerate below.

I will assume a flat $5 per participating pupil central administrative cost. In addition, I will consider costs associated with producing and distributing the examinations which serve as the basis of the performance assessment. The NSP has some experience with these costs and has found that production and distribution costs average $4.55 per participating pupil.[5]

Best Case Scenario

There are 7,500 participating students

7,500 * $5	$	37,500.
7,500 * $4.55	$	34,125.
Total	$	71,625.

Middle Case Scenario

There are 170,800 participating students

[5] During the pilot testing, the NSP spent $300,000 to produce and distribute exams for a total of 2,640 classes of students (660 in each of two subjects and 2 grade levels). If there are 25 pupils in each class, 66,000 students were involved. The per participating student cost is 300,000 / 66,000 = $4.55.

170,800	* $5	$ 854,000.
170,800	* $4.55	$ 777,140.
Total		$1,631,140.

Worst Case Scenario

There are 214,412 participating students

214,412	* $5	$1,072,060.
214,412	* $4.55	$ 975,575
Total		$2,047,635

III. Alternative Assumptions About the Absorption of Costs

I have now completed a set of estimates for the operations costs of a performance assessment system for a mid-size State where the assessment is focused on three grade levels in 2 subject areas. The costs I have totaled correspond to the dollar magnitudes of the ingredients that have been identified. No attention has been given to possible absorptions of costs through the displacement of existing practice. In this final section of

the cost analysis, I consider issues surrounding the possible absorption of the costs that have been enumerated. Recall that I dealt conceptually with this issue in Chapter 2. Once again, I make use of worst, middle, and best-case alternative scenarios.

In the following analyses, I make different assumptions about the magnitude of these absorptions. My rationale for doing so is that the assumed savings (however large they might be) occur because of the advent of performance assessment. Of course, it is possible that performance assessment occasions no savings or even generates additional costs at the local level. I explore the no savings result under the heading of the worst-case scenario. According to this view, performance assessment is a complete add-on and no local resources are released. I have not explored even more pessimistic scenarios, but the so-inclined reader is welcome to do so.

Within the middle and best-case scenarios I explore different views of how these savings could be realized. As the scenarios make clear, I see the potential for absorptions to arise in three areas: 1) local staff development; 2) the uses of class time for assessment (both preparation and the actual administration of the tasks); and 3) the utilization of assessment information.

(1) Local Staff Development

Worst Case Scenario

My presumption here is that resources currently being spent

at the local level on in-service staff development are productive and there is no potential for absorbing the costs of teachers acquiring the skills associated with performance assessment. Thus, there is no adjustment necessary to the costs.

Middle Case Scenario

Here my presumption is that local school districts will welcome opportunities to orient their teachers in the uses of performance assessment. It will be viewed as a substitution of a productive use of staff development resources for uses which were highly questionable in terms of their impact on teacher performance.

The willingness of local districts to make this substitution reduces the level of new resources that need to be devoted to teacher orientation. I assume further that these savings generate a 50% reduction in the costs associated with Scorer Training, Continuing Scorer Training, and Classroom Teacher Orientation.

The revised figures are:

Scorer Training

Best	6,263	*	.5	=	$3,132
Middle	717,360	*	.5	=	$358,680
Worst	2,358,532	*	.5	=	$1,179,266

Continuing Scorer Training

Best	1,875	*	.50	=	$938

Middle	59,621	*	.50	=	$29,811
Worst	300,300	*	.50	=	$150,150

Teacher Orientation

Best	16,250	*	.5	=	$8,125
Middle	1,802,435	*	.5	=	$901,218
Worst	4,762,976	*	.5	=	$2,381,488

Best Case Scenario

I assume here a 75% absorption. The revised figures for teacher orientation are:

Scorer Training

Best	6,263	*	.25	=	$1,566
Middle	717,360	*	.25	=	$179,340
Worst	2,358,532	*	.25	=	$589,633

Continuing Scorer Training

Best	1,875	*	.25	=	$469
Middle	59,621	*	.25	=	$14,905
Worst	300,300	*	.25	=	$75,075

Teacher Orientation

Best	16,250	* .25	=	$4,063
Middle	1,802,435	* .25	=	$450,609
Worst	4,762,976	* .25	=	$1,190,744

(2) The Use of Classroom Time for Performance Assessment

There are two issues here. First there is the degree to which time devoted within classrooms to performance assessment can function as time devoted simultaneously to instruction. However, even if the time devoted to performance assessment can function in this way, there is still a cost to consider because the allocated time comes at the expense of time previously committed to instruction. In other words, students as a consequence learn less of some things and more of other things as a result of the introduction of performance assessment (assuming the total amount of classroom time remains unchanged).

The second issue concerns the comparative productivity of the two instructional uses of classroom time. It is only to the degree that time devoted to performance assessment is a <u>more</u> productive instructional use of time than what was done previously with the time, that you find a local potential to absorb a portion of the classroom time costs of performance assessment.

In the worst, middle, and best case scenarios below, I explore the consequences surrounding different assumptions about the degree to which performance assessment uses of student time

are more productive than alternative uses.

Worst Case Scenario

Within this scenario performance assessment is an add-on to existing classroom activities. The implicit presumption is that the previous uses of classroom time are productive. This view does not deny that performance assessment time can have instructional benefits, but the view presumes that there is no potential for local levels to absorb or offset the costs.

Middle Case Scenario

Here I assume that 50% of the costs of classroom time devoted to both administration and preparation can be absorbed locally. The underlying view is that schools at present are spending resources in classrooms in rather unproductive ways so that it is a matter of doing fewer things that have little or no payoff in exchange for the opportunity to do more of something that has a good payoff.

The revised figures for Classroom time costs are:

Task Administration

Best	$72,000	*	.5	=	$36,000
Middle	$2,909,760	*	.5	=	$1,454,880
Worst	$3,415,680	*	.5	=	$1,707,840

Class Preparation

Best	$27,000	*	.5	=	$13,500
Middle	$2,182,320	*	.5	=	$1,091,160
Worst	$3,842,640	*	.5	=	$1,921,320

Best Case Scenario

Here I assume the relevant rate of absorption is 75%. The revised figures for classroom time costs are:

Task Administration

Best	$72,000	*	.25	=	$18,000
Middle	$2,909,760	*	.25	=	$727,440
Worst	$3,415,680	*	.25	=	$853,920

Class Preparation

Best	$27,000	*	.25	=	$6,750
Middle	$2,182,320	*	.25	=	$545,580
Worst	$3,842,640	*	.25	=	$960,660

(3) The Utilization of Assessment Information.

The central question here is the degree to which the new

assessment information actually makes a teacher's job easier. To the degree that the new information is easy to access and saves the teacher from devoting large amounts of time to pointless local testing activities, potentially large savings could be realized. These savings could even be larger than the cost of the time devoted to interpreting the results of the new assessments, thus giving rise to "negative costs."

Worst-Case Scenario.

No change is required here. The presumption is that there are no possible savings.

Middle Case Scenario

I assume a 50% rate of absorption.

The revised figures for the utilization of results are:

Best	$6,885	*	.5	=	$3,443
Middle	$613,830	*	.5	=	$306,915
Worst	$1,614,380	*	.5	=	$807,190

Best Case Scenario

The rate of absorption here is 75%. The revised figures are

as follows:

Best	$6,885	* .25	=	$1,721
Middle	$613,830	* .25	=	$153,458
Worst	$1,614,380	* .25	=	$403,595

IV. Summary

Table 5-1 provides a summary of the Operations Costs examined in this section. The table covers a two year period. Year 5 includes the costs of training the scorers as well as the Supplemental Lead Teachers. Year 6 is the first fully operational year of the project; my assumption is that in Year 6 no new Lead Teachers and no new scorers need to be trained. Scorer costs in Year 6 and beyond are limited to the estimated costs of maintaining an appropriately sized cohort of trained scorers (see the earlier section that deal with Continuing Scorer Training).

Table 5-1 About Here

According to my estimates, the operations costs in a mid-size State in Year 6 will range between a low of $210,000 and a high of $27.897 million. The middle case estimate is in the $9.291 to $14.967 million range, depending on how one wishes to treat the cost absorption issue. To place these numbers in some

perspective, if the pupil base for the state is 985,346 in grades K-12, the middle case operations costs expressed on a per pupil basis range between 15.19 and 9.43 dollars.

Table 5-2 summarizes the operations cost totals from Table 5-1 on a per pupil basis (using the 985,346 pupil count).

Table 5-2

Summary of Operations Costs in Year 6

For a Mid-Size State with 985,346 Pupils

	Worst	Middle	Best
Best	.31	.25	.21
Middle	15.19	11.35	9.43
Worst	28.31	21.24	17.71

Note: The column headings refer to assumptions about the degree of cost absorption; the row headings refer to assumptions about the magnitude of program required to achieve the intended results. Cell entries are $/pupil.

CHAPTER 6

OPERATIONS COSTS IN A SMALL STATE

I. Introduction

The focus in this chapter is on a small state where I assume there are 268 elementary schools and 48 secondary schools organized into 40 local education agencies (i.e., school districts). I will assume further that the state's grade 4 enrollment is 7,256 and that the enrollments in grades 8 and 10 are 6,799 and 5,958, respectively.

I deal explicitly with the following components of operations costs: (1) Supplemental Lead Teacher Training; (2) Scorer Training; (3) Continuing Scorer Training; (4) Outside Auditing; (5) Administration of Tasks; (6) Scoring; (7) Utilization of Results; and (8) Administration and Overhead (including the costs of printing and distributing the exams). Next I consider alternative assumptions regarding the possible absorptions of selected cost components. The chapter concludes with an overview and summary of the cost estimates.

II. Components of Cost

(1) Supplemental Lead Teacher Training

Available Supply of Lead Teachers

Recall that as a by-product of the Development Phase there is a pool of trained scorers. Under the assumptions I imposed in Chapter 3, I estimated the size of this pool per year to be 1,300 trained scorers per year (400 trained at the national scoring meetings and 900 trained regionally). At the end of the 4 year Development Phase, the maximum number of trained scorers will be 1,300 * 4 or 5,200, assuming there are no scorers who repeat their training program. This also presumes that there is no loss of skill over as long as 4 years for teachers who learn to score at the outset of the project.

Given the likelihood that scorers will vary in how well they learn the requisite skills, that some decay will take place over time for those who are trained early, and that the project will lose track of some participants, I will make alternative assumptions regarding the actual size of the reservoir of scorers that is available at the end of the Operations Phase.

According to the best-case scenario, there is little loss over time and teachers learn the relevant skills quite easily and uniformly. In other words, NSP does not have to deal with significant unevenness in how well teachers learn to be scorers. Nor is there much unevenness in how well the trained teachers retain their skills. Nor does the project lose track of many scorers over time. The middle and worst-case scenarios relax these assumptions and introduce potentially significant levels of

unevenness, depreciation and obsolescence, and loss.

There are no obvious benchmarks to rely upon in assigning magnitudes to the discount factors that need to be used, so I will make the relatively arbitrary assumption that under the best-case scenario, the effective loss is 10%, and that under the middle and worst-case scenarios, the respective percentage losses are 20 and 30.

My assumption is that these experienced scorers constitute the initial NSP representation in the field. These people will play lead roles in the training and implementation of the project within the participating states. They will be involved in both the performance assessment as well as the cumulative portfolio development aspects of the NSP. I will refer to them as Lead Teachers.[1]

I also assume that these Lead Teachers are divided across the participating states in proportion to the respective states' populations. My rationale for this is based on the NSP practice of varying the number of invitations to the national scoring meetings according to its partners' populations (recall that either 2, 3, or 4 teachers from each grade level and subject were invited).

---

[1] According to NSP documentation, performance tasks constitute just one part of the cumulative portfolio that will be generated for each student. While there will be central guidance provided about the types of items that should be included in students' portfolios, much discretion will be maintained at the individual school and teacher levels. The Lead Teachers will provide training and assistance to front-line teachers who are participating in the project.

Recall that this is a small state with 268 elementary schools and 48 secondary schools with a total pupil enrollment in grades 4, 8, and 10 of 20,013. I assume that this state received 2 * 2 * 3 or 12 nationally trained scorers each year (during the Development Phase), and that the number of regional training workshops that were conducted within the state is proportional to the state's share of the NSP base student population (i.e., the population from all the participating states and units). According to NSP documentation, a state with 20,013 pupils would comprise 0.4 per cent of the pupil base being served by the project. Thus, I assume that it operated 0.4 per cent of the 30 regional workshops that were held each year. This corresponds to .12 workshop per year, but I will also assume that each state operates at least one workshop.[2] Recall that each workshop generated 30 trained scorers. It follows that for each year the small State created a pool of 1.00 * 30 or 30 locally trained scorers. This yields a total of 12 + 30.0 or 42 potential Lead Teachers each year for a total possible of 168. The application of the best, middle, and worst case loss rates that I derived above generates the following estimates of Lead Teacher Supply for the small State at the close of the Development Phase of the Project:

---

[2] Strictly speaking I should make a downward adjustment for the remaining states given this decision to assume that each state offers at least one regional scoring workshop. Also, note that I have been willing to accept fractions of workshops when the number is greater than unity.

Best Case

$168 - (.1 * 168) = 151.20$

Middle Case

$168 - (.2 * 168) = 134.40$

Worst Case

$168 - (.3 * 168) = 117.60$

Demand for Lead Teachers During Operations

The next question is whether the supply of these Lead Teachers thanks to the Development Phase of the Project will be adequate to staff the Operations Phase of the Project. To begin to answer this question, I make a series of assumptions about the scope of the operational phase of the performance assessment project. There are two dimensions to this demand: (1) the number of schools that will be involved in the operational version of performance assessment; and (2) the level of direct supervision by Lead Teachers that is required within each participating school. I will be making alternative assumptions regarding each dimension, and I shall join them within the scenario framework so that the best case of one is linked with the best case of the other. In other words, I will not be considering alternative combinations of best, middle, and worst cases along each dimension.

Counts of Participating Schools

At one extreme, I will assume that in order for the project

to achieve its goals, it will be necessary to implement an annual performance assessment program within every school in each participating state. I call this a census approach to implementation, and it corresponds to a worst-case scenario with respect to the associated costs.

Recall that a major goal of the NSP is to change fundamentally the conduct of instruction throughout entire schooling systems. According to this worst-case cost scenario, it is necessary to have an NSP presence within every school during every year of the operations phase of the project. I also make alternative assumptions about the level of the presence that is required, but for now the focus is on how many schools need to participate in a given year during the operational phase.

At the other extreme, I will assume that it is possible for the project to achieve its goals through the use of a light matrix sampling design. The presumption here will be that a periodic program of assessment within a relatively small sample of schools is sufficient within each state to achieve the far-reaching goals of the NSP. The sample of schools and classrooms participating will vary from year to year. All schools and the relevant classrooms will be eligible for selection, and at any given time teachers and administrators will not know when their classrooms and schools will participate. Moreover, in any given year, I assume that the state will focus on some subset of the possible tasks within the Task Bank. This scenario will correspond to a best-case view of costs since fewer resources will be required (by

assumption) for the project to achieve its goals.

The middle case scenario involves a situation where there is interest in district specific results. In contrast to the census and matrix approaches, the presumption here is that there is interest in district level performance. The design will require sampling from within districts and this will require a measure of performance assessment that lies between the first two extremes that I have identified.

Level of Direct Supervision Provided By Lead Teachers

According to the NSP proposal, a goal of the project is to have two externally trained and certified scorers within each school participating in the performance assessment activities. I am assuming that such people correspond to what I have called Lead Teachers, and I note that there is some ambiguity surrounding the precise level of Lead Teacher supervision that will be appropriate. At one extreme, it could be that two Lead Teachers could handle all of the testing taking place within a school regardless of the subject being taught. Thus, in a secondary school with grades 8 and 10 present, two Lead Teachers could handle the testing program for both mathematics and language arts. From a cost perspective, this extreme corresponds to a best-case scenario.

At the opposite extreme, it may be necessary to have two Lead Teachers for each grade and subject being assessed. In this case,

a secondary school with two grade levels would require 8 Lead Teachers. This reality corresponds to a worst-case scenario in terms of costs.

A middle ground can be defined by thinking of the Lead Teachers as being able to cross grade levels but not subject areas. Under the terms of this middle-case scenario, the secondary school with grades 8 and 10 would require 4 Lead Teachers.

These three scenarios (for both the number of participating schools and the number of needed Lead Teachers in each school) are used below to define the demand for Lead Teachers in a typical operational year of the project.

Best Case Scenario

Number of participating schools. This scenario involves the use of a matrix sampling design. I assume that the sampling goal will be 100 observations per task,[3] and that in any given year the State will employ 25 per cent of the tasks available within the Task Bank. If every student participating in the program received one task, implementation would require 25 * 100 or 2,500 pupils per grade per subject. But, I will also assume that each student participating will complete 2 tasks, and thereby reduce the

[3] According to the NSP, a sample of 2,500 observations needs to be drawn for all 25 tasks for each subject at each grade level being considered to satisfy psychometric concerns over validity.

required number of participating students by half to 1,250 per grade per subject.

Recall that the mid-size State has a grade 4 enrollment of 7,256 and a population of 268 elementary schools. If the grade 4 students are evenly distributed across the schools, it follows that the average school will enroll approximately 27 4th grade students. If the goal is to have 2,500 participating 4th grade students (1,250 per subject), in a given year performance assessment will need to take place within approximately 93 of the 268 elementary schools (2,500/27).

At the secondary level, there are two grade levels. For the small State, there are 6,799 8th grade students and 5,958 10th grade students. There are 48 secondary schools and assuming all 8th grade students are enrolled in secondary schools and that the students are evenly distributed across the schools, it follows that each school enrolls, on average, approximately 133 students at each of these grade levels. If the goal is to have 2,500 participating 8th and 2,500 participating 10th grade students (again, 1,250 per subject), then in a given year performance assessments will need to take place within approximately 19 of the 48 secondary schools (2,500/133).

Level of staffing. In keeping with the best-case scenario, I assume that each participating school needs 2 Lead Teachers and that these Lead Teachers can handle both multiple subject areas and grade levels (where they occur). If there are 93 elementary schools in the program, there will need to be (93*2) or 186 Lead

Teachers for the elementary schools. If there are 19 secondary schools in the program, there will need to be (19*2) or 38 Lead Teachers for the secondary schools.

Thus, the best-case scenario involves a total annual demand of 186+38=224 Lead Teachers. This compares with the derived supply of approximately 151. Thus, even under terms of the best-case scenario, the small State will need to provide training for an additional 73 Lead Teachers.

Implications for costs. I assume the necessary training will take the form of a supplemental series of 4 day workshops structured around scoring exercises. Recall that these workshops cost $18,810 and yield 30 trained scorers. Thus, the supplemental cost for Lead Teacher training for the small State will be:

((224 - 151) / 30) * $18,810 = $45,771.

Middle Case Scenario

Number of participating schools. Here the idea is that the state is interested in having information from each district, and the presumption is that the matrix sampling design described above misses a significant number of districts. As I indicated earlier, the mid-size State operates 40 separate school districts. I assume that the average grade 4 enrollment within each district is 7,256 / 40 or 181 and that the average grade 8 and 10 enrollments are 6,799 / 40 or 170 and 5,958 / 40 or 149, respectively. Using

the 27 4th grade pupils per school and 133 8th or 10th grade pupils per school figures that I derived above, it follows that on average each district operates 6.7 elementary schools and 1.21 secondary schools.

I assume that a sample of 3 elementary schools per district and 1 secondary school per district will be adequate to provide the district level aggregates.[4] This will require staffing performance assessment activities in 120 elementary schools and 40 secondary schools in a given year.

Level of staffing. In accordance with the middle-case scenario where the assumption is that 2 Lead Teachers are needed for each subject within each school, there is an implied demand 120 * 2 * 2 = 480 Lead Teachers for the elementary program and 40 * 2 * 2 = 160 Lead Teachers for the secondary program.[5] This means the state needs a pool of 640 Lead Teachers compared to the 134 that are available following the Development Phase.

Implications for costs. Again, I assume the necessary training will take the form of a supplemental series of 4 day workshops structured around scoring exercises. The same costs that I derived earlier will apply. Recall that these workshops

---

[4] For both the large and mid-size States, I assumed a sample of 2 elementary schools per district. I choose 3 here because the small State operates a relatively large number of quite small elementary schools.

[5] Recall that the assumption is that Lead Teachers can cross grade levels but not subject areas. This explains why the secondary schools require 4 rather than 8 Lead Teachers.

cost $18,810 and yielded 30 trained scorers. Thus, the supplemental cost for Lead Teacher training for the mid-size State according to the middle case scenario will be:

((640 - 134) / 30) * $18,810 = $317,262

Worst Case Scenario

Number of participating schools. Recall that the worst-case scenario involves a census approach to performance assessment where the presumption is that every 4th, 8th, and 10th grade student needs to be assessed every year in both subject areas.

The small State has a population of 268 elementary and 48 secondary schools. If the state pursues a census approach, Lead Teacher staffing will be required in each of these schools.

Level of staffing. According to the worst-case scenario, 2 Lead Teachers are needed for each possible combination of subject and grade level. Assuming elementary schools involve only grade 4, the total number of Lead Teachers needed for the 4th grade assessment program will be 268 * 2 * 2 or 1,072. The corresponding number of Lead Teachers for the 8th and 10th grade assessment programs (assuming they are all located within the secondary schools) will be 48 * 2 * 2 * 2 or 384.

Implications for costs. The total number of Lead Teachers needed according to this scenario is 1,072 + 384 = 1,456. In contrast, according to the worst-case scenario, the Development

Phase of the project generates a supply of 118 Lead Teachers. The relevant cost calculation (assuming the Lead Teachers are trained through the use of regional workshops) is:

((1,456 - 118) / 30) * $18,810 = $838,926.

(2) Scorer Training

Best Case Scenario

Number of Scorers Needed

Assuming each participating student generates 2 tasks, the annual total number of tasks that need to be scored will be the number of students per grade level (2,500) * the number of grade levels (3) * the number of tasks completed (2) = 15,000.

I assume that each scorer scores 400 tasks. This is the equivalent of 8 days of work. The NSP does not seek to develop a supply of "professional" task scorers. It is, instead, committed to achieving a broad base of participation among teachers and others. For this reason, I impose the 400 task ceiling.

If there are 15,000 tasks to score in the large State, and if each scorer scores 400, the demand for scorers will be 37.5.

Level of Training Required

Minimal training will be required to train local scorers under terms of the best-case scenario. The underlying assumption is that this kind of assessment and its scoring will be very much in-line with how teachers think and go about their work. The teachers are presumed to adapt quickly and easily. I assume that the training can be done quite informally within the local districts; as a consequence travel costs for participants become negligible and will be omitted. Since the Lead Teachers will be traveling, I have included an allowance for this travel in the budget.

However, for the sake of deriving cost estimates, I will continue to treat the training as if it has a group workshop nature. In particular, I will assume that what is necessary is the equivalent of a one-day workshop for 30 participants where the participant/Lead Teacher ratio is 8:1.

$100 per diem for 1 day for 30 participants	3,000.
travel:$20 average for 30 participants	n.a.
lodging:	n.a.
food and misc $10/day per participant	300.
materials (20/participant)	600.
Lead Teacher costs	
assuming 3.7 Lead Teachers per 30 participants	
$250 per diem per day per Lead Teacher	925.
travel costs at $40 per	

Lead Teacher	148.
lodging	n.a.
food and misc.: $10/day per instructor	37.
Total Cost per 1 Day Scoring Workshop	$5,010.

The yield for this workshop is 30 trained scorers. If the need for scorers is 37.5, the costs of training these individuals will be $6,263.

Middle Case Scenario

Number of Scorers Needed

According to this scenario, there will be assessment activities in 120 elementary schools and 40 high schools. The average number of 4th grade students per elementary school is 27; the average number of 8th and 10th grade students is 133. Thus, there are 13,880 students participating in a given year. If each student completes 4 tasks (two for each of two subjects), there will be 55,520 tasks to score.

If scorers score 400 tasks each, there will be a demand for 139 scorers.

Level of Training Required

A more ambitious level of training is required under the terms of the middle-case scenario. Instead of the equivalent of a 1 day (30 person) scoring workshop, I will assume that 2 days are necessary. I shall also assume that a more intensive training experience is necessary. Instead of the 8:1 ratio of participants to Lead Teachers, I will assume that a 4:1 ratio is necessary. I shall also build travel costs into the budget, since my presumption is that it will be less possible for the training to place informally at the home sites.

$100 per diem for 2 days for 30 participants	6,000.
travel:$20 average for 30 participants (per day)	1,200.
lodging:	n.a.
food and misc $10/day per participant	600.
materials (20/participant)	600.
Lead Teacher costs	
assuming 7.5 Lead Teachers per 30 participants	
$250 per diem per day per Lead Teacher	3,750.
travel costs at $20 per day per instructor	300.
lodging	n.a.
food and misc.: $10/day per instructor	150.
Total Cost per 2 Day Scoring Workshop	$12,600.

Assuming there are 139 scorers that need to be trained, and assuming these 2 day workshops each yield 30 trained scorers, the cost of scorer training will be $58,380.

Worst Case Scenario

Number of Scorers Needed

The mid-sized State operates 268 elementary schools and 48 secondary schools. Assuming there are 27 4th grade students per elementary school and 133 8th and 10th grade students per secondary school, there will be 20,004 students being assessed each year. If each student completes 4 tasks, there will be 80,016 tasks to score each year.

Assuming each scorer scores 400 tasks, there will be a demand for 200 scorers.

Level of Training Required

Since this is the worst case scenario, I assume that teachers, on balance, find it difficult to grasp the requisite skills to function effectively as scorers. I assume that these teachers need to spend the equivalent of 4 one-day workshops acquiring these skills, and that these workshops will be offered regionally. I assume further that the scorer/Lead Teacher ratio

in the workshop needs to be 2:1.

For each 4 one-day elementary task scoring workshop, there will be the following costs:

$100 per diem for 4 days for 30 participants	12,000.
travel:$20 average for 30 participants (per day)	2,400.
lodging:	n.a.
food and misc $10/day per participant	1,200.
materials (20/participant)	600.
Lead Teacher costs	
assuming 15 Lead Teachers per 30 participants	
$250 per diem per day per Lead Teacher	15,000.
travel costs at $20 per day per Lead Teacher	1,200.
lodging	n.a.
food and misc.: $10/day per Lead Teacher	600.
Total Cost per 2 Day Scoring Workshop	$33,000.

According to the worst case scenario, there will be a need for 200 trained scorers. Assuming these workshops produce 30 trained scorers, the cost of developing this network of scorer support will be $220,000.

(3) Continuing Scorer Training

Best Case Scenario

According to this scenario, teachers find scoring to be a quite enjoyable and professionally enriching activity. They actively seek opportunities to learn how to do it, and once employed only rarely give up the job voluntarily. Moreover, there is considerable cross-over from the old tasks to the new so that there is a minimal need for formal retraining of those who continue.

To operationalize this view of the reality, I assume that what is required is the equivalent of 1/2 a day of a scorer's time to meet with a group of fellow scorers to discuss their activities. I envision a series of very small informal workshops where groups of scorers essentially teach and refresh themselves.

Cost of the 1/2 day 30 participant local district workshop:

$100 perdiem for 1/2 day per scorer	50
travel:	n.a.
lodging:	n.a.
food and misc	n.a.

Total Number of Scorers according to the Best-Case Scenario: 37.5

Total Cost for Continuing Scorer Development:

37.5 * $50 = $1,875

Middle Case Scenario

If the middle-case scenario is accurate, there will be a moderate degree of turnover among scorers. Teachers are presumed to find scoring an interesting but demanding activity. It is presumed to be viewed positively but as a burden that needs to be shared equitably. Also, some degree of carry-over will be presumed to exist between old and new tasks, so that the teachers remaining as scorers require only modest amounts of new training.

I operationalize this scenario by assuming that the recurring training needs amount to a one-day 30 participant regional workshop for 1/5 of the scoring cohort each year. The workshop will be taught by Lead Teachers and the ratio of participants to Lead Teachers will be 8:1.

Cost of a one-day regional workshop

$100 perdiem for 1 day for 30 participants	3,000.
travel:$20 average for 30 participants (per day)	600.
lodging:	n.a.
materials ($10/participant)	300

food and misc $10/day per participant	300.
Lead Teacher costs	
assuming 3.7 instructors per 30 participant workshop	
$250 perdiem per Lead Teacher	925.
travel costs at $20 per day per instructor	74.
lodging	n.a.
food and misc.: $10/day per instructor	37.
Cost per 1 Day Continuing Staff Development Workshop	$5,236.

Total number of scorers
in the middle-case scenario = 139

Total Cost for Continuing Staff Development:

(((.2) * (139)) / 30) * $5,236 = $4,852

Worst Case Scenario.

According to the worst case scenario, teachers will find scoring quite burdensome. They will avoid having to perform the service and they will seek to quit the job at the first opportunity. Thus, whatever efficiencies are gained thanks to experience will be lost because of the resulting high level of turnover. The high turnover will generate large and continuing demands for scorer training.

Moreover, this scenario holds that there will be little

carry-over from prowess as a scorer with one set of tasks to performance as a scorer on new tasks that are developed. Thus, even those remaining on the job will need periodic training.

I assume that within this scenario, a training program for 1/3 of the scorer cohort will be required, on average, each year. This program will be divided into training for both new scorers who replace those exiting the system and "refresher-type" training for those who are continuing.

I assume that the magnitude of this program will correspond to the cost of a 2 full day regional workshop organized for 30 participants. I also assume that the Lead Teachers will serve as instructors and that the participant/Lead Teacher ratio will be 4:1. The costs of such a workshop are these:

Costs per 2 day Continuing Staff Development Workshop:

$100 per diem for 2 days for 30 participants	6,000.
travel:$20 average for 30 participants (per day)	1,200.
lodging:	n.a.
food and misc $10/day per participant	600.
materials ($20/participant)	600
Lead Teacher Costs	
assuming 7.5 Lead Teachers per 30 participants	
$250 perdiem per Lead Teacher	3,750.
travel costs at $20 per day	

per Lead Teacher	300.
lodging	n.a.
food and misc.: $10/day per Lead Teacher	150.
Total Cost per 2 Day Continuing Staff Development Workshop	12,600.

Total Number of Participants: (200) / 3 = 67

Total Cost of Providing Continuing Staff Development:

(67 / 30) * $12,600 = $28,140.

(4) Outside Auditing

I assume that one of the by-products of difficulty teaching teachers how to score will be a need for outside auditing; the greater the difficulty, the greater the need for outside auditing. The necessary auditing will not be confined to the performance tasks; the cumulative portfolios will also be subject to periodic audit.

Best Case Scenario

Here I assume that the Lead Teachers themselves can handle whatever auditing needs to be done. I also assume that they can do this during the equivalent of 1 full day per year. The implicit presumption is that the system works quite well and that only periodic spot checks are necessary. The Lead Teachers would

not audit their own schools.

Cost of a 1 day block of time for 1 Auditor to Work

$250 per diem	250.
travel ($20/day)	20.
meals, etc. ($10/ day)	10.
lodging	n.a.
Total	280.

Recall that the number of Lead Teachers within the Best-Case Scenario is 224. If all of the Lead Teachers participate in the auditing phase of the project, the cost will be 224 * $280 = $62,720.

Middle Case Scenario

Within this scenario, auditing is a more serious problem. Again, I assume that all of the Lead Teachers are involved and that they need to meet and work the equivalent of 2 full days each.

Cost of a 2 day block of time for 1 Auditor to Work

$250 per diem (2 days)	500.
travel ($20/day)	40.

meals, etc. ($10/ day)	20.
lodging	n.a.
Total	$560.

Recall that the number of Lead Teachers according to the Middle-Case Scenario is 640. If the cost of the program is $560 per Lead Teacher and there are 640 Lead Teachers, the cost will be $358,400.

Worst Case Scenario

By assumption, the costs incurred to provide a relatively large amount of intensive training will not be sufficient to offset the difficulties teachers encounter as they seek to develop their scoring skills. I assume the training reduces but does not eliminate the problem. The failure to solve the problem through training necessitates the installation of a relatively extensive auditing system which will involve outside scorers routinely reviewing the performance exams and cumulative portfolios produced throughout the system. Double scoring will be commonplace. Perhaps even triple scoring.

Moreover, the public relations problems could be immense, particularly if the auditors are systematically lowering scores for a school, or if high stakes begin to be attached to these scores. These public relations needs can generate significant

additional costs, but I will make no attempt here to estimate their magnitudes.

I continue to assume that the Lead Teachers can perform the auditing work but that they will each require the equivalent of 4 full days to accomplish their goals. I also assume that this work will require periodic regional meetings and therefore generates travel costs.

Cost of a 4 day period of time for 1 Auditor to Work

$250 perdiem	1,000.
travel ($20/day)	80.
meals, etc. ($10/ day)	40.
lodging	n.a.
Total	1,120.

Recall that the number of Lead Teachers provided for within the worst-case scenario is 1,456. This implies an auditing cost of $1,630,720.

(5) Administration of Tasks

I have divided this section into two portions: A) Teacher Orientation, and B) Classroom Implementation. The Classroom Implementation section is also divided into two portions: 1) class time devoted to actual assessment, and 2) class time devoted to

preparation.

A. TEACHER ORIENTATION

Best Case Scenario

My assumption here is that teachers will respond to the performance assessment approach quite readily. A 1/2 day 30 participant orientation program at the local level for all teachers that will be administering tasks and assembling cumulative student portfolios is all that is required. Note, however, that I am not dealing with whatever orientation might be necessary for teachers who are not directly involved in the administration of the exams (i.e., those at grade levels other than 4, 8, and 10). I assume a 30:1 ratio of participants to Lead Teachers. I also assume that these meetings will take place regionally. I have not provided an allowance for substitute teacher costs on the grounds that if the workshop takes place during regular school hours, the stipend paid to the teachers would logically be used to compensate the substitute teacher who is covering the teacher's class.

$100 per diem for .5 day	
for 30 participants	1,500.
travel:	n.a.
lodging:	n.a.

food and misc $10/day per participant	n.a.
instructor costs	
assuming 1 instructor per 30 participant workshop	
$250 per diem per day per instructor	125.
travel costs at $20 per day per instructor	n.a.
lodging	n.a.
food and misc.: $10/day per instructor	n.a.
Total Cost per .5 Day Teacher Orientation Workshop for 30	$1,625.

The number of teachers requiring this orientation in a given year corresponds to the number of participating classrooms. Recall that under the terms of the best-case scenario, a total of 2,500 pupils will be assessed at each grade level each year (1,250 in each subject). This yields a total of 7,500 pupils. If there are 25 pupils in each class, this corresponds to a count of 300 classroom teachers. Assuming it costs $1,625 to orient a group of 30 teachers, the total cost of orientation will be $16,250.

Middle Case Scenario

Here I assume that the teacher orientation is less easily accomplished. In particular, I assume that the program requires the equivalent of a 1 day 30 participant regional workshop where the participant Lead Teacher ratio is 15:1.

The cost of such a workshop will be:

$100 per diem for 1 day	
for 30 participants	3,000.
travel:$20 average for 30 participants	600
lodging:	n.a.
food and misc $10/day per participant	300.
Lead Teacher Costs	
assuming 2 Lead Teachers per 30	
participants	
$250 per diem per day per	
Lead Teacher	500.
travel costs at $20 per day	
per Lead Teacher	40.
lodging	n.a.
food and misc.: $10/day	
per Lead Teacher	20.
Total Cost per 1 Day Teacher Orientation	
Workshop (for 30 participants)	$4,460.

The number of teachers requiring this orientation can be derived from the number of students being assessed under the terms of the middle-case scenario. These counts are: 3,240 4th grade students, 5,320 8th grade students, and 5,320 10th grade students. Assuming 25 students to a class and assuming the participating 4th grade students are being assessed in both subjects by the same teacher, the number of 4th grade teachers requiring orientation will be 3,240/25 = 130. At the 8th and 10th grade levels, it is likely that participating students will be taught by two different teachers. Thus, if there are 5,320 8th grade students and if they are being assessed in 2 subject areas by different teachers and if the relevant pupil-teacher ratio is 25, the number of 8th grade teachers needing orientation will be (5,320/25) * 2 = 426. Similarly, 426 10th grade teachers will need to be oriented, according to this scenario.
It follows that the total number of teachers requiring orientation will be 982.

If the cost for orienting 30 teachers is $4,460, the cost of orienting this number of teachers is (982/30) * $4,460 = $145,991.

Worst Case Scenario

Here the Lead Teachers fail in their effort to convey enthusiasm about performance assessment to their colleagues. Front-line teachers view performance assessment as a burden imposed on them by external authorities, and the Lead Teachers

have no choice but to make a relatively intensive effort to orient teachers.

I assume that this translates into a need to provide the equivalent of a 2 full day workshop for every participating teacher. Note: While I am costing this orientation in terms of a formal workshop, the reality is likely to be quite different with Lead Teachers working individually with front-line teachers.

I calculate the costs of mounting an orientation program for these teachers on the assumption that the workshop will be delivered regionally to groups of 30 teachers and that the relevant participant/instructor ratio is 7.5:1

Cost of a 2 day regional workshop for 30 participants

$100 per diem for 2 days for 30 participants	6,000.
travel:$20 average for 30 participants	1,200.
lodging:	n.a.
food and misc $10/day per participant	600.
Lead Teacher Costs	
assuming 4 Lead Teachers for 30 participants	
$250 per diem per day per Lead Teacher	2,000.
travel costs at $20 per day	

per Lead Teacher	160.
lodging	n.a.
food and misc.: $10/day	
per Lead Teacher	80.
Total Cost per 2 Day Teacher Orientation Workshop	10,040.

According to the Worst Case scenario, 7,236 4th grade, 6,384 8th grade, and 6,384 10th grade students need to be assessed. Again, assuming that the relevant pupil-teacher ratio is 25 and that the 4th grade teachers handle 2 subjects, 289 4th grade teachers will need orientation. With the same pupil-teacher ratio and assuming each teacher handles 1 subject at the 8th and 10th grade levels, 511 8th and 511 10th grade teachers will need orientation. The total number of teachers needing orientation is 1,309.

If it costs $10,040 to orient 30 teachers, then the total cost of teacher orientation will be (1,309/30) * $10,040 = $438,079.

B. CLASSROOM IMPLEMENTATION

1)) Class Time Devoted to Actual Assessment

The assumption I impose here is that the amount of time teachers spend actually administering performance tasks will be

the same regardless of whether it is a worst, middle, or best case scenario. For each grade level and subject, I assume that each task on average requires a total of 3 class hours. I also assume that over the course of a year a student will complete 2 tasks.

Thus, for each class participating in the assessment, the time required will be 6 hours. In addition, I will assume that the teacher must spend 1 hour in preparation for each task. It follows that the teacher preparation time will be 2 hours, aside from the time spent being oriented.

The next step is to figure the cost, on average, of an hour of class time. I assume that the cost of an hour of teacher time is $25, and I adjust this figure upward by $5 to account for miscellaneous costs such as space, materials, utilities, and administrative overhead. Students' time is clearly required for the administration of performance assessment tasks, but there is no satisfactory means of recognizing its value in these cost calculations. For now, I note that students' time has value and is required by performance assessment activities, but I do not attempt to include estimates of its value in these cost calculations.

According to the best-case scenario, there will be 300 teachers that need to be oriented. This figure gives us a basis for assuming that the number of classes that will be involved in a given year will be 300. The corresponding figures for the middle-case and worst-case scenarios are: 982 and 1,309, respectively. Thus, the costs of actually administering the performance tasks

are:

Best Case	300	*	8	*	$30	=	$72,000.
Middle Case	982	*	8	*	$30	=	$235,680.
Worst Case	1,309	*	8	*	$30	=	$314,160.

2)) Class Time Devoted to Preparation

I assume that teachers take time from instruction to prepare their students for performance assessments. Again, there is a question about whether such time can simultaneously serve an instructional purpose, and I deal with this issue later in the treatment of cost absorption. For now I treat preparation time as a cost and I use the best, middle, and worst case scenarios to examine varying assumptions about how much time is devoted on average by teachers to preparation.

Best Case Scenario

Here my assumption is that .5 hour of preparation is spent for each 1 hour of class time devoted to performance assessment. The cost is: 300 * 6 * 0.5 * $30 = $27,000.

Middle Case Scenario

I assume here that 1.0 hours of preparation accompanies each hour of time devoted to performance assessment. Under this assumption, the costs of time devoted tc class preparation will be: 982 * 6 * 1.0 * $30 = $176,760

Worst Case Scenario

I assume that for each hour of performance assessment, teachers within this scenario devote 1.5 hours of class time to preparation.[6] Under this assumption, the costs of time devoted to preparation will be:

1,309 * 6 * 1.5 * $30 = $353,430.

(6) Scoring

Best Case Scenario

Recall that there will be 15,000 tasks to score each year under the terms of the best-case scenario. There are 37.5 trained scorers in place, each handling 400 tasks. This requires 8 full days of work (50 tasks per day for 8 days). And I will assume that these scorers will be paid a stipend of $250 per day for this

---

6 The 3:1 ratio between the best and worst case scenarios is not entirely arbitrary. The Office of Technology Assessment (1992, pg. 29) found that teachers in a large urban school district reported devoting up to 3 hours of preparation for each test administration. I am taking the upper figure here to reflect the worst-case scenario in terms of costs.

work.

Total scoring cost will be 37.5 * 8 * $250 = $75,000

Middle Case Scenario

According to this scenario, there will be 139 scorers working 8 days at $250 per day. This yields a total scoring cost of 139 * 8 * $250 = $278,000.

Worst Case Scenario

The worst case (census approach) requires 200 scorers. The associated costs are: 200 * 8 * $250 = $400,000.

(7) Utilization of Results

It is important to include estimates of the costs associated with making use of the performance assessment results. I estimate these costs by making alternative assumptions about how much teacher time and Lead Teacher time will be required per hour of classroom time devoted to performance assessment.

Best Case Scenario

Here the teachers adapt quite readily to the use of

performance assessment results. They require minimal supervision from Lead Teachers. I assume that for every hour of class time devoted to performance assessment a teacher requires .12 of an hour of his/her time studying the results. I also assume that for every hour a classroom teacher devotes to reflecting on performance assessment results, .06 of an hour of Lead Teacher time will be required. This will be time spent working primarily one-to-one with the classroom teachers interpreting results and providing guidance.

Under these assumptions the costs of utilizing the results of performance assessment will be:

300 classes * 6 hours = 1,800 class-hours
1,800 class-hours * .12 = 216 additional teacher hours
216 * $30 = $6,480

In addition, the costs of the Lead Teachers' time need to be added.

216 * .06 = 12.96 hours

Assuming 8 hour days, this translates into 1.62 work days for Lead Teachers. Assuming their daily rate is $250, this involves an additional $405.

Total Cost = $6,480 + $405 = $6,885.

Middle Case Scenario

Here I assume that the teachers need to spend .25 hours for every hour of class time devoted to performance assessment, and that the Lead Teachers need to spend .12 of an hour for each teacher-hour devoted to interpretation.

982 classes * 6 hours = 5,892 class-hours
5,892 class-hours * .25 = 1,473 additional teacher hours
1,473 * $30 = $44,190.

In addition, the costs of the Lead Teachers' time need to be added.

1,473 * .12 = 177 hours.

Assuming 8 hour days, this translates into 22 days of work for Lead Teachers. Assuming their daily rate is $250, this involves an additional $5,500.

Total Cost = $44,190 + $5,500 = $49,690.

Worst Case Scenario

Here teachers, on average, require considerable instruction

and supervision in the utilization of performance assessment results. I assume that for every hour devoted to performance assessment a teacher requires .5 of an hour of his/her time studying the results. I also assume that the for every hour a classroom teacher spends interpreting test results a Lead Teacher needs to spend .25 hours.

1,309 classes * 6 hours = 7,854 class-hours
7,854 class-hours * .50 = 3,927 additional teacher hours
3,927 * $30 = $117,810.

In addition, the costs of the Lead Teachers' time need to be added.

3,927 * .25 = 982 hours

Assuming 8 hour days, this translates into 123 days of work for Lead Teachers. Assuming their daily rate is $250, this involves an additional $30,750.

Total Cost = $117,810 + $30,750 = $148,560.

(8) Administration and Overhead

There will be central administrative costs at both the national and individual state levels. The national costs will need to be spread across the various participating states and

units. For now, I will conceive of central administrative support as a State level matter. Contributions to the national level will be made out of the costs I enumerate below.

I will assume a flat $5 per participating pupil central administrative cost. In addition, I will consider costs associated with producing and distributing the examinations which serve as the basis of the performance assessment. The NSP has some experience with these costs and has found that production and distribution costs average $4.55 per participating pupil.[7]

Best Case Scenario

There are 7,500 participating students

7,500	*	$5	$ 37,500.
7,500	*	$4.55	$ 34,125.
Total			$ 71,625.

Middle Case Scenario

There are 13,880 participating students

---

[7] During the pilot testing, the NSP spent $300,000 to produce and distribute exams for a total of 2,640 classes of students (660 in each of two subjects and 2 grade levels). If there are 25 pupils in each class, 66,000 students were involved. The per participating student cost is 300,000 / 66,000 = $4.55.

13,880	*	$5	$ 69,400.
13,880	*	$4.55	$ 63,154.
Total			$ 132,554.

Worst Case Scenario

There are 20,004 participating students

20,004	*	$5	$ 100,020.
20,004	*	$4.55	$ 91,018
Total			$ 191,038

III. Alternative Assumptions About the Absorption of Costs

I have now completed a set of estimates for the operations costs of a performance assessment system for a small State where the assessment is focused on three grade levels in 2 subject areas. The costs I have totaled correspond to the dollar magnitudes of the ingredients that have been identified. No attention has been given to possible absorptions of costs through

the displacement of existing practice. In this final section of the cost analysis, I consider issues surrounding the possible absorption of the costs that have been enumerated. Recall that I dealt conceptually with this issue in Chapter 2. Once again, I make use of worst, middle, and best-case alternative scenarios.

In the following analyses, I make different assumptions about the magnitude of these absorptions. My rationale for doing so is that the assumed savings (however large they might be) occur because of the advent of performance assessment. Of course, it is possible that performance assessment occasions no savings or even generates additional costs at the local level. I explore the no savings result under the heading of the worst-case scenario. According to this view, performance assessment is a complete add-on and no local resources are released. I have not explored even more pessimistic scenarios, but the so-inclined reader is welcome to do so.

Within the middle and best-case scenarios I explore different views of how these savings could be realized. As the scenarios make clear, I see the potential for absorptions to arise in three areas: 1) local staff development; 2) the uses of class time for assessment (both preparation and the actual administration of the tasks); and 3) the utilization of assessment information.

(1) Local Staff Development

Worst Case Scenario

My presumption here is that resources currently being spent at the local level on in-service staff development are productive and there is no potential for absorbing the costs of teachers acquiring the skills associated with performance assessment. Thus, there is no adjustment necessary to the costs.

Middle Case Scenario

Here my presumption is that local school districts will welcome opportunities to orient their teachers in the uses of performance assessment. It will be viewed as a substitution of a productive use of staff development resources for uses which were highly questionable in terms of their impact on teacher performance.

The willingness of local districts to make this substitution reduces the level of new resources that need to be devoted to teacher orientation. I assume further that these savings generate a 50% reduction in the costs associated with Scorer Training, Continuing Scorer Training, and Classroom Teacher Orientation.

The revised figures are:

Scorer Training

Best	6,263	*	.5	=	$3,132
Middle	58,380	*	.5	=	$29,190
Worst	220,000	*	.5	=	$110,000

Continuing Scorer Training

Best	1,875	*	.50	=	$938
Middle	4,852	*	.50	=	$2,426
Worst	28,140	*	.50	=	$14,070

Teacher Orientation

Best	16,250	*	.5	=	$8,125
Middle	145,991	*	.5	=	$72,996
Worst	438,079	*	.5	=	$219,040

Best Case Scenario

I assume here a 75% absorption. The revised figures for teacher orientation are:

Scorer Training

Best	6,263	*	.25	=	$1,566
Middle	58,380	*	.25	=	$14,595
Worst	220,000	*	.25	=	$55,000

Continuing Scorer Training

Best	1,875	*	.25	=	$469
Middle	4,852	*	.25	=	$1,213

Worst	28,140	*	.25	=	$7,035

Teacher Orientation

Best	16,250	*	.25	=	$4,063
Middle	145,991	*	.25	=	$36,498
Worst	438,079	*	.25	=	$109,520

(2) The Use of Classroom Time for Performance Assessment

There are two issues here. First there is the degree to which time devoted within classrooms to performance assessment can function as time devoted simultaneously to instruction. However, even if the time devoted to performance assessment can function in this way, there is still a cost to consider because the allocated time comes at the expense of time previously committed to instruction. In other words, students as a consequence learn less of some things and more of other things as a result of the introduction of performance assessment (assuming the total amount of classroom time remains unchanged).

The second issue concerns the comparative productivity of the two instructional uses of classroom time. It is only to the degree that time devoted to performance assessment is a <u>more</u> productive instructional use of time than what was done previously with the time, that you find a local potential to absorb a portion

of the classroom time costs of performance assessment.

In the worst, middle, and best case scenarios below, I explore the consequences surrounding different assumptions about the degree to which performance assessment uses of student time are more productive than alternative uses.

Worst Case Scenario

Within this scenario performance assessment is an add-on to existing classroom activities. The implicit presumption is that the previous uses of classroom time are productive. This view does not deny that performance assessment time can have instructional benefits, but the view presumes that there is no potential for local levels to absorb or offset the costs.

Middle Case Scenario

Here I assume that 50% of the costs of classroom time devoted to both administration and preparation can be absorbed locally. The underlying view is that schools at present are spending resources in classrooms in rather unproductive ways so that it is a matter of doing fewer things that have little or no payoff in exchange for the opportunity to do more of something that has a good payoff.

The revised figures for Classroom time costs are:

Task Administration

Best	$72,000	*	.5	=	$36,000
Middle	$235,680	*	.5	=	$117,840
Worst	$314,160	*	.5	=	$157,080

Class Preparation

Best	$27,000	*	.5	=	$13,500
Middle	$176,760	*	.5	=	$88,380
Worst	$353,430	*	.5	=	$176,715

Best Case Scenario

Here I assume the relevant rate of absorption is 75%. The revised figures for classroom time costs are:

Task Administration

Best	$72,000	*	.25	=	$18,000
Middle	$235,680	*	.25	=	$58,920
Worst	$314,160	*	.25	=	$78,540

Class Preparation

Best	$27,000	*	.25	=	$6,750
Middle	$176,760	*	.25	=	$44,190

Worst	$353,430	*	.25	=	$88,358

(3) The Utilization of Assessment Information.

The central question here is the degree to which the new assessment information actually makes a teacher's job easier. To the degree that the new information is easy to access and saves the teacher from devoting large amounts of time to pointless local testing activities, potentially large savings could be realized. These savings could even be larger than the cost of the time devoted to interpreting the results of the new assessments, thus giving rise to "negative costs."

Worst-Case Scenario.

No change is required here. The presumption is that there are no possible savings.

Middle Case Scenario

I assume a 50% rate of absorption.

The revised figures for the utilization of results are:

Best	$6,885	*	.5	=	$3,443
Middle	$49,690	*	.5	=	$24,845

Worst	$148,560	*	.5	=	$74,280

Best Case Scenario

The rate of absorption here is 75%. The revised figures are as follows:

Best	$6,885	*	.25	=	$1,721
Middle	$49,690	*	.25	=	$12,423
Worst	$148,560	*	.25	=	$37,140

IV. Summary

Table 6-1 provides a summary of the Operations Costs examined in this section. The table covers a two year period. Year 5 includes the costs of training the scorers as well as the Supplemental Lead Teachers. Year 6 is the first fully operational year of the project; my assumption is that in Year 6 no new Lead Teachers and no new scorers need to be trained. Scorer costs in Year 6 and beyond are limited to the estimated costs of maintaining an appropriately sized cohort of trained scorers (see the earlier section that deal with Continuing Scorer Training).

Table 6-1 About Here

According to my estimates, the operations costs in a small State in Year 6 will range between a low of $241,000 and a high of

$3.504 million. The middle case estimate is in the $921,000 to $1.383 million range, depending on how one wishes to treat the cost absorption issue. To place these numbers in some perspective, if the pupil base for the state is 94,779 in grades K-12, the middle case operations costs expressed on a per pupil basis range between 14.59 and 9.72 dollars.

Table 6-2 summarizes the operations cost totals from Table 6-1 on a per pupil basis (using the 94,779 pupil count).

Table 6-2

Summary of Operations Costs in Year 6
For a Small State with 94,779 Pupils

	Worst	Middle	Best
Best	3.52	2.87	2.54
Middle	14.59	11.34	9.72
Worst	36.97	30.21	26.83

Note: The column headings refer to assumptions about the degree of cost absorption; the row headings refer to assumptions about the magnitude of program required to achieve the intended results. Cell entries are $/pupil.

CHAPTER 7

OVERVIEW AND DISCUSSION

I. Introduction

The cost estimates provided in Chapters 3-6 need to be interpreted carefully. They are, as I have indicated repeatedly, heavily dependent on assumptions; moreover, the alternative assumptions give rise to wide discrepancies across the estimates. Such wide discrepancies across cost estimates are not unusual in this type of inquiry. Haney, Madaus, & Lyons (1993, p. 118), for example, estimated that the total investment in state and district testing programs currently is between $311 million and $22.7 billion dollars, annually. In light of these wide discrepancies, policymakers need to exercise restraint and avoid choosing those estimates that most closely accord with points of view chosen for perhaps quite unrelated reasons. This chapter draws the high and low estimates together and stresses the importance of viewing the results collectively.

One byproduct of contrasting the estimates from each of the three prototypical states is insight into the likely nature of scale economies in the development of this kind of reform. Recall that the size of the three states varies widely. Because the numerous assumptions and caveats apply equally across the three

states, it is meaningful to compare the resulting cost estimates head-to-head, and this chapter considers the policy implications of these comparisons.

As the discussion in Chapter 2 attempted to make clear, it is much more problematic to compare the cost estimates generated in Chapters 3-6 with the cost estimates for different types of pupil assessment programs. Such comparisons, strictly speaking, require controls for differences in the nature and magnitude of the benefits being generated.

However, it does not follow that the magnitudes I have estimated need to be viewed only in isolation. It can be meaningful to place these estimates in context of other existing or anticipated expenditures of resources, so long as differences in the anticipated benefits are kept in sight. In this chapter, I put my cost estimates into context by comparing them with resource commitments in a number of different areas including: state spending on K-12 education; Federal spending on technical assistance centers; estimates of total spending on teacher inservice training; and alternative estimates of resources required for national pupil assessment programs. The chapter closes with a brief overview of the uses of cost analysis in public policy development and implementation.

II. Comparing Costs Across the Different Sized States

Table 7-1 is based on the operations cost estimates reported

in Chapters 4-6, expressed on a per pupil basis. The table is designed to illustrate the potential for economies of scale to arise in the implementation of pupil performance assessment and the conduct of systemic reform initiatives.

Table 7-1

Unit Cost Differences of Performance Assessment
Across Different Sized States
(Assuming No Cost Absorption)*

	Large Size	Mid-Size
Mid-Size		
Best	3.4	
Middle	1.3	
Worst	1.0	
Small Size		
Best	39.1	11.4
Middle	1.2	1.0
Worst	1.3	1.3

*Cell entries are the unit costs associated with the state's size indicated by the row heading divided by the unit cost associated with the state's size indicated by the column heading. For example, the 3.4 at the top of the first column indicates that the unit cost of the best case scenario in the mid-size State is 3.4 times larger than the unit cost of the best case scenario for the large State. Changes in the assumption about the magnitude of cost absorptions do not alter the results reported in this table.

Table 7-1 reports evidence of scale economies. As a rule, the cell entries are greater than 1.0, and this suggests that unit costs are higher in smaller compared to larger states. The table also suggests that the most pronounced scale economies appear under terms of the best case scenario. The cell entries are clearly larger for all of the best case rows. This result is a byproduct of the lack of sensitivity between the terms of the best case scenarios sampling requirements and the size of the state. Recall that the same number of students needed to be sampled regardless of the size of the state. Under these circumstances, the larger the state, the larger the number of students over which the costs of the fixed size program can be spread. Hence, there arise considerable scale economies.

Finally, Table 7-1 also suggests that scale economies are more pronounced between the small size and the mid-size states than between the mid-size and the large size states. This is the case for both the best and the worst case scenarios. An exception occurs for the middle case scenario where the unit costs are roughly the same between the small and the middle size states and then drop for the large size state.

These results suggest that, ceteris paribus, small states will find it more burdensome to implement pupil performance assessment and systemic reform initiatives than will larger states. These extra costs may be offset to the degree that collaboration can occur across state boundaries. There may also be a useful role for the Federal government to play at equalizing

the opportunities across states to participate in these reform initiatives.

III. Placing Operations Cost Estimates into Context

Relative to total state spending levels on K-12 schools. My worst case operations cost estimate for the large State is $97.386 million, or .7% of what a state of comparable size spent on its public elementary and secondary schools in 1991.[1] The comparable percentages for the mid-size and small States are .6% and .7%, respectively.[2] It therefore appears that the resource requirements of the kind of systemic reform I envision in my interpretation of the NSP amount to less than 1% of the resources currently being devoted by these states to elementary and secondary public schools. Moreover, these percentages correspond to the resource requirements of the worst-case scenario that I envisioned. Not only do the percentages reflect the worst case scenario, they are also calculated under the assumption that none of the costs were absorbed at the local level.

---

[1] The state I am using here for comparison purposes is Texas where current spending on public elementary and secondary education was $13,444 million in 1991 (U.S. Bureau of the Census 1992).

[2] The mid-size comparison state I used was Virginia where current spending on public elementary and secondary schools in 1991 was $4,996 million. The comparison small state was Vermont where current spending in 1991 on public elementary and secondary schools was $507 million (U.S. Bureau of the Census 1992).

Relative to Federal spending on technical assistance. According to estimates provided by the U.S. Department of Education, Federal support for Technical Assistance Centers in FY 1993 amounted to $53.5 million. These Assistance Centers, in part, provide staff development services for teachers. My large, mid, and small State prototypes represent roughly 8, 2.5 and .2 per cent of the nation's pupil population.[3] If I apportion the $53.5 million across the states in proportion to these pupil population figures, I obtain the following distribution:

large state (8%) $4.28 million
mid-sized state (2.5%) $1.34 million
small state (.2%) $0.107 million

In contrast, Table 7-2 summarizes my estimates of the resources that will be devoted to Year 6 staff development.

---

3 U.S. Bureau of the Census (1992, p. 149.)

Table 7-2

Summary of Staff Development Costs in Year 6 of Operations*

	Degree of Absorption		
	0%	50%	75%
Large State			
Best	.025	.012	.006
Middle	5.715	2.858	1.429
Worst	21.887	10.944	5.471
Mid-Size State			
Best	.025	.012	.006
Middle	2.476	1.238	.619
Worst	6.677	3.338	1.670
Small State			
Best	.025	.012	.006
Middle	.201	.100	.049
Worst	.615	.307	.154

*This table is based on figures drawn from Tables 4-1, 5-1, and 6-1. I have included the following categories of cost in these summations: Supplemental Lead Teacher Training, Scorer Training, Continuing Scorer Training, Teacher Orientation, and Utilization of Results. Note, however, that the summations are from Year 6 where there are no anticipated costs for both Supplemental Lead Teacher Training and Scorer Training.

It is clear that Federal spending on Technical Assistance Centers corresponds roughly to the resource requirements of the middle case scenario that I envision for staff development.

Relative to total spending on staff development. According to a recent study of Chapter 1 implementation supported by the U.S. Department of Education, the annual average number of days devoted in 1991-92 to staff development for U.S. elementary school teachers is approximately 3 days. The corresponding figure for secondary teachers is 2.2 days (Millsap, Moss and Gamse 1993, p.7-

2).

During 1990 there were an estimated 1,379 thousand elementary teachers and 1,012 thousand secondary teachers (U.S. Bureau of the Census 1992, p. 148). If I apply my 8%, 2.5%, and .2% adjustments to transform the national figures into estimates for my large, mid-size, and small states, I obtain the following estimates for the three states:

	Elementary	Secondary
Large	110,320	80,960
Mid-Size	34,475	25,300
Small	2,758	2,024

If we use the 8 hour day and $25 per hour conventions that I adopted for the previous chapters, a first approximation of the resources currently being devoted to staff development (but with no allowance for the costs of providing inservice programs) within each of the 3 states I am considering is:

(in millions)

	Elementary	Secondary	Total
Large	$66.192	$35.622	$101.814
Mid-Size	$20.685	$11.132	$ 31.817
Small	$ 1.655	$ .891	$ 2.546

These figures are clearly much larger than even the worst

case scenario figures I report in Table 7-2. Keep in mind, however, that these figures pertain to all teachers, not just those participating in the kind of systemic reform efforts envisioned by the NSP.

Relative to previous estimates of national testing costs. Perhaps the closest previous attempt to estimate the costs of performance assessment was conducted by the U.S. General Accounting Office (GAO). The GAO reported that estimates of the cost of a national testing system have ranged from a few million dollars a year up to $3 billion (U.S. GAO, p.2). The GAO based their estimates on a 1991 survey of testing officials in all state education agencies plus a random sample of U.S. school districts. On the basis of these survey results, the GAO estimated that the overall cost of systemwide testing in 1990-91 was $516 million.

The GAO identified 3 testing models, no one of which corresponds perfectly with the reforms envisioned within the NSP proposal. However, the model coming closest to the NSP formulation involves a decentralized system of clusters of states where each cluster uses a different performance based test. A crucial difference between what the GAO envisions and the NSP involves the use of cumulative portfolios and their periodic assessment. According to the GAO, the decentralized system would be the most expensive of the models they considered, and would cost on the order of $330 million per year for operations, nationwide.

As I pointed out earlier, my large, mid, and small state

prototypes represent roughly 8, 2.5 and .2 per cent of the nation's population. If I apportion the $330 million across the states in proportion to these population figures, I obtain the following distribution:

large state:	$26.4 million
mid-size state:	$ 8.25 million
small state	$ 0.660 million

If we choose the middle case scenario and the worst case assumption about cost absorption, the comparable figures are:

large state	$39.0 million
mid-size state:	$15.0 million
small state:	$ 1.4 million

Keep in mind, however, that these figures are not strictly comparable. The NSP program includes a sizeable investment in staff development while the GAO estimates are based primarily on the costs of administering the exams. The NSP program is a more comprehensive reform designed to foster a systemic change in how education is practiced nationwide; the GAO estimates are of a nationwide testing program that is much less ambitious in its scope.

It is worth noting that the GAO report is one of the few published studies where Development Costs of a national testing

system are considered separately. Their estimate is $100 million. This corresponds to my best, middle, and worse case scenario estimates of $15.85 million, $16.34 million, and $16.96 million, respectively, where I am simply taking the sum of my year by year estimates over the 4 year development period (see Chapter 3). In other words, I have not discounted future costs, largely because it is unclear how the GAO report handled costs over time, and it seems best to keep the analysis simple.

Thus, my Development Cost estimates are significantly below the GAO estimates, but keep in mind that the program I envision retains development activities during operations.

IV. Concluding Comments

I have sought in this monograph to generate upper and lower bounds on the likely resource requirements of the kind of systemic reform envisioned within the NSP. This is, of necessity, a delicate exercise, since the project itself is still being developed and implementation is on-going. It has been particularly difficult to generate cost estimates for the cumulative portfolio aspects of the NSP. These initiatives are quite new, and their design is evolving. Some preliminary findings suggest that teachers find themselves devoting considerable amounts of time to the portfolios, both in class and outside of class (Koretz, Stcher, & Deibert (1992). As we gain experience with the portfolio component of systemic reform,

significant adjustments may be necessary in the cost estimates I have generated here. In contrast, more is known about the resources required to develop formal assessment tasks, and greater confidence can surround the associated cost estimates.

Systemic reform, as it is currently understood, most certainly does not lend itself to a conventional cost analysis, not to mention cost-effectiveness or cost-benefit analysis. Nevertheless, policymakers need guidance about what resource requirements are likely to be, and the kind of cost analysis I have presented here is intended to provide this assistance.

I shall refrain from drawing conclusions about the costs being high or low in any sort of global comparative sense. Indeed, I have gone about as far as I can prudently go by placing my cost estimates "in context." I shall also refrain from offering guidance about what strikes me as the best combination of scenarios and assumptions about the appropriate level of cost absorption. Policymakers are in a better position to make these judgments, since they will or should have some vision of the scale of the enterprise they seek to establish.

In closing, I want to re-emphasize the importance of being attentive to the cost dimension of policymaking. As difficult and limited as cost analyses tend to be, I am thoroughly convinced that their neglect places policymakers on a direct path toward poor results and the worst kinds of unpleasant surprises during implementation.

REFERENCES

American Association for the Advancement of Science, (1989). Science for all Americans. (AAAS Project 2061 overview report). Washington, D.C.: Author.

Barnett, W. Steven. (1985). "Benefit-Cost Analysis of the Perry Preschool Program and its Policy Implications," Educational Evaluation and Policy Analysis 7: 333-342.

Barnett, W. Steven (1991). "Benefits of Compensatory Preschool Education," Journal of Human Resources XXVII,2: 279-312.

Barro, Stephen (1991) "Cost of Education Differentials Across the States" typescript, U.S. Department of Education, NCES.

Bauer, Ernest A. (1992). "NATD Survey of Testing Practices and Issues," Educational Measurement: Issues and Practice 11,1 (Spring): 10-14.

Bowman, M.J. (1966). "The Costing of Human Resource Development," in E.A.E. Robinson and J. Vaizey, eds., The Economics of Education. London: Macmillan.

Buchanan, J.M. (1966). Cost and Choice. Chicago: Markham.

Catterall, James S. (1987). "On the Social Costs of Dropping Out of School," The High School Journal (Oct./Nov.): 19-30.

Catterall, James S. (1988). "Estimating the Costs and Benefits of Large-Scale Assessments: Lessons From Recent Research," paper presented at the annual conference of the American Educational Research Association.

Chubb, John E. and Terry M. Moe (1990). Politics, Markets, and America's Schools. Washington, D.C., The Brookings Institution, Basic Books.

Clune, William H. (1993). "Systemic Educational Policy: A Conceptual Framework," in Susan H. Fuhrman (ed.), Designing Coherent Education Policy. San Francisco, CA: Jossey-Bass Publishers, pp. 125-140.

Fowler, William J., Jr. (1992) "What Should We Know About School Finance," paper presented at the annual conference of the American Education Finance Association, New Orleans, La.

Furhman, Susan H. (ed.) (1993). Designing Coherent Education Policy. San Francisco, CA: Jossey-Bass Publishers.

Gong, Brian, Colleen Lahart, and Rosalea Courtney (1991). "Current State Science Assessments: Is Nothing Bettern Than Something?" Educational Testing Service.

Haller, Emil J. (1974). "Cost Analysis for Educational Program Evaluation," in W. James Popham (ed.), Evaluation in Education: Current Applications. Berkeley, CA: McCutchan Publishing Co.: 401-450.

Haney, Walter M., George F. Madaus and Robert Lyons (1993). The Fractured Marketplace for Standardized Testing. Boston: Kluwer Academic Publishers.

Jamison, D.T., S.J. Klees, and S.J. Wells (1978). The Costs of Educational Media. Beverly Hills, CA: Sage Publications.

Koretz, Daniel, Brian Stecher, and Edward Deibert (1992). "The Vermont Portfolio Assessment Program: Interim Report on Implementation and IMpact, 1991-1992 School Year," The RAND Corporation.

Levin, Henry M. (1983). Cost-Effectiveness: A Primer. Beverly Hills, California: Sage Publications.

Levin, Henry M. (1991). "Cost-Effectiveness at Quarter Century," in McLaughlin, M. and Phillips (eds.), Evaluation and Education: At Quarter Century. Chicago: University of Chicago Press: 189-209.

Levin, Henry M., Gene V. Glass, and Gail R. Meister (1984). "Cost-Effeciveness of Four Educational Interventions," Project Report No. 84-A11, Institute for Research on Educational Finance and Governance, School of Education, Stanford University.

Marshall, Ray and Marc Tucker (1992). Thinking for a Living: Education and the Wealth of Nations. New York: Basic Books.

Millsap, Mary Ann, Marc Moss and Beth Gamse (1993). The Chapter 1 Implementation Study: Final Report, Chapter 1 in the Public Schools. U.S. Department of Education, Office of Policy and Planning, Washington, D.C.

Monk, David H. (1982). "Alternative Perceptions of Cost and the Resource Allocation Behaviors of Teachers," Educational Administration Quarterly 18,2 (Spring): 60-80.

Monk, David H. (1990). Educational Finance: An Economic Approach. (New York: McGraw-Hill).

Monk, David H. (1992). "Education Productivity Research: An Update and Assessment of Its Role in Education Finance Reform," Educational Evaluation and Policy Analysis, 14,4: 307-332.

Monk, David H. and Jennifer A. King (in press), " Cost Analysis as a Tool for Education Reform," in Stephen L. Jacobson and Roabert Berne (eds.), The Reform of Education. Newbury Park, CA: Corwin Press.

National Center for Education Statistics (1991). Digest of Education. Washington, D.C.: U.S. Department of Education, Author.

New Standards Project (1992). "The New Standards Project, 1992-1995, A Proposal," Rochester, N.Y.: National Center on Education and the Economy and Pittsburgh, Pa.: Learning Research and Development Center, Author.

O'Day, Jennifer A. and Marshall S. Smith (1993). "Systemic Reform and Educational Opportunity," in Susan H. Fuhrman (ed.), Designing Coherent Education Policy. San Francisco, CA: Jossey-Bass Publishers, pp. 250-312.

Pelavin Associates (1992). "Evaluating Education Reform: Assessment of Student Performance," author.

Shepard, Lorrie A. and Amelia E. Kreitzer, (1987), "The Texas Teacher Test," Educational Researcher 16,6: 22-31.

Thomas, Hywel (1990). Education Costs and Performance: A Cost-Effectiveness Analysis. London: Cassell Educational Limited.

U.S. Bureau of the Census, (1992). Statistical Abstract of the United States. U.S. Department of Commerce, Washington, D.C.

U.S. Congress, Office of Technology Assessment, (1992). Testing in American Schools: Asking the Right Questions, OTA-SET-519 (Washington, D.C.: U.S. Government Printing Office, February)

U.S. General Accounting Office (1993). Student Testing: Current Extent and Expenditures, With Cost Estimates for a National Examination, GAO/PEMD-93-8 (Washington, D.C.: U.S. General Accounting Office, Program Evaluation and Methodology Division).

Walsh, V.C. (1970). Introduction to Contemporary Microeconomics. New York: McGraw-Hill.

Table 3-4
Task Development Costs

			TABLE 3-4		
			Task Development Costs		
		Year 1	Year 2	Year 3	Year 4
Number of Required					
Raw Tasks					
In Mathematics					
		[1]			
Best		150.00	100.00	88.20	83.30
Middle		150.00	107.10	100.00	93.80
Worst		150.00	125.00	115.40	115.40
Average Unit Costs					
		[2]			
Best		1,800	1,600	1,400	1,200
Middle		1,900	1,800	1,700	1,600
Worst		2,000	1,900	1,800	1,700
Total Costs					
For Mathematics Tasks					
Best		270,000	160,000	123,480	99,960
Middle		285,000	192,780	170,000	150,080
Worst		300,000	237,500	207,720	196,180
Total Costs					
For Language Arts Tasks					
Best		540,000	256,000	160,524	99,960
Middle		570,000	347,004	272,000	210,112
Worst		600,000	475,000	415,440	392,360
Total Costs for					
Math and Lang Arts					
Task Development					
Best		810,000	416,000	284,004	199,920
Middle		855,000	539,784	442,000	360,192
Worst		900,000	712,500	623,160	588,540

[1] **Note:** The need is for 25 usable tasks for each of 3 grade levels. In Year 1, for all 3 scenarios, the assumed rate of loss is 50%. Thus, the entry is 25x3x2=150.

[2] **Note:** The relevant ratio of the 2 unit prices is 80/20. Thus, the cell entry is .8(2000)+.2(1000)=1800.

Table 3-6
Summary of Development Costs
(in Millions of Dollars)

		Year 1	Year 2	Year 3	Year 4	Year 5
Administrative Overhead		0.08	0.08	0.08	0.08	
Task Development						
Best		0.81	0.42	0.28	0.2	
Middle		0.86	0.54	0.44	0.36	
Worst		0.9	0.71	0.62	0.59	
Task Refinement		0.65	0.65	0.65	0.65	
Test Product and Dist.		0.45	0.45	0.45	0.45	
Administ of Pilot Test		0.792	0.792	0.792	0.792	
Task Calibration		0.15	0.15	0.15	0.15	
Scoring		1.367	1.367	1.367	1.367	
Interpret of Results		0.045	0.045	0.045	0.045	
Continuing Task Devel						
Best						0.027
Middle						0.048
Worst						0.079
Total						
Best		4.344	3.954	3.814	3.734	0.027
Middle		4.394	4.074	3.974	3.894	0.048
Worst		4.434	4.244	4.154	4.124	0.079

Table 4-1
Operations Costs in a
Large State

	YEAR 5			YEAR 6		
	Worst Case	Mid-Case	Best Case	Worst Case	Mid-Case	Best Case
Ajustments on Items with **	0%	50%	75%	0%	50%	75%
Supplemental Lead Teacher Training						
Best	0	0	0	0	0	0
Middle	7.209	7.209	7.209	0	0	0
Worst	16.934	16.934	16.934	0	0	0
Scorer Training**						
Best	0.006	0.003	0.002	0	0	0
Middle	1.907	0.953	0.477	0	0	0
Worst	7.809	3.904	1.952	0	0	0
Continuing Scorer Training**						
Best	0	0	0	0.002	0.001	0
Middle	0	0	0	0.158	0.079	0.04
Worst	0	0	0	0.994	0.497	0.248
Outside Auditing						
Best	0.031	0.031	0.031	0.031	0.031	0.031
Middle	6.72	6.72	6.72	6.72	6.72	6.72
Worst	30.742	30.742	30.742	30.742	30.742	30.742
Admin of Tasks						
Teacher Orient.**						
Best	0.016	0.008	0.004	0.016	0.008	0.004
Middle	4.662	1.985	0.992	4.662	1.985	0.992
Worst	15.604	7.802	3.901	15.604	7.802	3.901
Classroom Implemt						
Time for Tests**						
Best	0.072	0.036	0.018	0.072	0.036	0.018
Middle	7.526	3.763	1.882	7.526	3.763	1.882
Worst	11.19	5.595	2.798	11.19	5.595	2.798
Time for Prep**						
Best	0.027	0.014	0.007	0.027	0.014	0.007
Middle	5.645	2.822	1.411	5.645	2.822	1.411
Worst	12.589	6.295	3.147	12.589	6.295	3.147
Scoring						
Best	0.075	0.075	0.075	0.075	0.075	0.075
Middle	9.08	9.08	9.08	9.08	9.08	9.08
Worst	14.198	14.198	14.198	14.198	14.198	14.198

(in Millions of Dollars)

Table 4-1
Operations Costs in a
Large State

Utilization of Results**							
Best		0.007	0.003	0.002	0.007	0.003	0.002
Middle		1.588	0.794	0.397	1.588	0.794	0.397
Worst		5.289	2.645	1.322	5.289	2.645	1.322
Administ Overhead							
Best		0.072	0.072	0.072	0.072	0.072	0.072
Middle		4.336	4.336	4.336	4.336	4.336	4.336
Worst		6.78	6.78	6.78	6.78	6.78	6.78
Totals							
Best		0.306	0.242	0.211	0.302	0.24	0.209
Middle		48.673	37.662	32.504	39.715	29.579	24.858
Worst		121.135	94.895	81.774	97.386	74.554	63.136

(in Millions of Dollars)

Table 5-1
Operations Costs in a
Mid-Size State

	YEAR 5			YEAR 6		
	Worst Case	Mid-Case	Best Case	Worst Case	Mid-Case	Best Case
Ajustments on Items with **	0%	50%	75%	0%	50%	75%
Supplemental Lead Teacher Training						
Best	0	0	0	0	0	0
Middle	2.518	2.518	2.518	0	0	0
Worst	5.128	5.128	5.128	0	0	0
Scorer Training**						
Best	0.006	0.003	0.002	0	0	0
Middle	0.717	0.359	0.179	0	0	0
Worst	2.359	1.179	0.59	0	0	0
Continuing Scorer Training**						
Best	0	0	0	0.002	0.001	0
Middle	0	0	0	0.06	0.03	0.015
Worst	0	0	0	0.3	0.15	0.075
Outside Auditing						
Best	0.032	0.032	0.032	0.032	0.032	0.032
Middle	2.352	2.352	2.352	2.352	2.352	2.352
Worst	9.341	9.341	9.341	9.341	9.341	9.341
Admin of Tasks						
Teacher Orient.**						
Best	0.016	0.008	0.004	0.016	0.008	0.004
Middle	1.802	0.901	0.451	1.802	0.901	0.451
Worst	4.763	2.381	1.191	4.763	2.381	1.191
Classroom Implemt						
Time for Tests**						
Best	0.072	0.036	0.018	0.072	0.036	0.018
Middle	2.91	1.455	0.727	2.91	1.455	0.727
Worst	3.416	1.708	0.854	3.416	1.708	0.854
Time for Prep**						
Best	0.027	0.014	0.007	0.027	0.014	0.007
Middle	2.182	1.091	0.546	2.182	1.091	0.546
Worst	3.843	1.921	0.961	3.843	1.921	0.961
Scoring						
Best	0.075	0.075	0.075	0.075	0.075	0.075
Middle	3.416	3.416	3.416	3.416	3.416	3.416
Worst	4.288	4.288	4.288	4.288	4.288	4.288

(in Millions of Dollars)

Table 5-1
Operations Costs in a
Mid-Size State

Utilization of Results**							
Best		0.007	0.003	0.002	0.007	0.003	0.002
Middle		0.614	0.307	0.153	0.614	0.307	0.153
Worst		1.614	0.807	0.404	1.614	0.807	0.404
Administ Overhead							
Best		0.072	0.072	0.072	0.072	0.072	0.072
Middle		1.631	1.631	1.631	1.631	1.631	1.631
Worst		2.048	2.048	2.048	2.048	2.048	2.048
Totals							
Best		0.307	0.243	0.212	0.303	0.241	0.21
Middle		18.142	14.03	11.973	14.967	11.183	9.291
Worst		36.8	28.801	24.805	29.613	22.644	19.162

(in Millions of Dollars)

Table 6-1
Operations Costs in a
Small State

	YEAR 5			YEAR 6		
Worst Case	Mid-Case	Best Case	Worst Case	Mid-Case	Best Case	
Ajustments on Items with **	0%	50%	75%	0%	50%	75%
Supplemental Lead Teacher Training						
Best	0.046	0.046	0.046	0	0	0
Middle	0.317	0.317	0.317	0	0	0
Worst	0.839	0.839	0.839	0	0	0
Scorer Training**						
Best	0.006	0.003	0.002	0	0	0
Middle	0.058	0.029	0.015	0	0	0
Worst	0.22	0.11	0.055	0	0	0
Continuing Scorer Training**						
Best	0	0	0	0.002	0.001	0
Middle	0	0	0	0.005	0.002	0.001
Worst	0	0	0	0.028	0.014	0.007
Outside Auditing						
Best	0.063	0.063	0.063	0.063	0.063	0.063
Middle	0.358	0.358	0.358	0.358	0.358	0.358
Worst	1.631	1.631	1.631	1.631	1.631	1.631
Admin of Tasks						
Teacher Orient.**						
Best	0.016	0.008	0.004	0.016	0.008	0.004
Middle	0.146	0.073	0.036	0.146	0.073	0.036
Worst	0.438	0.219	0.11	0.438	0.219	0.11
Classroom Implemt						
Time for Tests**						
Best	0.072	0.036	0.018	0.072	0.036	0.018
Middle	0.236	0.118	0.059	0.236	0.118	0.059
Worst	0.314	0.157	0.079	0.314	0.157	0.079
Time for Prep**						
Best	0.027	0.014	0.007	0.027	0.014	0.007
Middle	0.177	0.088	0.044	0.177	0.088	0.044
Worst	0.353	0.177	0.088	0.353	0.177	0.088
Scoring						
Best	0.075	0.075	0.075	0.075	0.075	0.075
Middle	0.278	0.278	0.278	0.278	0.278	0.278
Worst	0.4	0.4	0.4	0.4	0.4	0.4

(in Millions of Dollars)

Table 6-1
Operations Costs in a
Small State

Utilization of Results**							
Best		0.007	0.003	0.002	0.007	0.003	0.002
Middle		0.05	0.025	0.012	0.05	0.025	0.012
Worst		0.149	0.074	0.037	0.149	0.074	0.037
Administ Overhead							
Best		0.072	0.072	0.072	0.072	0.072	0.072
Middle		0.133	0.133	0.133	0.133	0.133	0.133
Worst		0.191	0.191	0.191	0.191	0.191	0.191
Totals							
Best		0.384	0.32	0.289	0.334	0.272	0.241
Middle		1.753	1.419	1.252	1.383	1.075	0.921
Worst		4.535	3.798	3.43	3.504	2.863	2.543

(in Millions of Dollars)