IMPLEMENTING CURRICULAR CHANGE
THROUGH STATE-MANDATED TESTING:
ETHICAL ISSUES

SANDRA MATHISON, State University of New York at Albany

Many states use standardized testing to control what happens in schools. Tests determine who can teach, who can be promoted to the next grade, and who can graduate with what sort of credentials. The current public dissatisfaction with schooling has left state offices of education casting about for any way to improve education. According to one common view, teachers do not teach what they should—the state curriculum or syllabus. State education authorities have several mechanisms at their disposal for controlling curriculum—for example, statewide textbook adoption in Texas. Porter has analyzed the various forms external standards can take in the pursuit of good schooling.¹ In New York, and I suspect in many other states, state-mandated testing exerts a major control over what is taught and how. Cohen has viewed assessment as an information system out of which arises policy initiatives.² Madaus, however, views assessment instrumentally as the means by which policy initiatives are implemented.³ Testing has spread to all levels of schooling and to most subjects, serving a multiplicity of purposes—accountability, student assessment, student certification, and program evaluation.

The logic is simple: Testing certain content in certain ways will result in an alignment of classroom practices with the official view of what and how subjects should be taught—what Popham calls a "curricular magnet."⁴ In the last decade, the use of state-mandated tests has become a predominate strategy in educational reform efforts, and no state has escaped the appeal of this

popular means for improving education. So, for example, the recently implemented New York state-mandated 4th grade science program evaluation test will ensure that science is taught, and taught in a certain way. The same is true in social studies, which is now tested at 6th and 8th grades in New York State. Although implementing these tests may cause great anxiety for students and school personnel and, in fact, may be inefficient in attaining the intended outcomes, the state still uses the policy-implementation strategy. The end justifies the means, even when the means do not produce the desired ends perfectly.

The effects of testing are for the most part clearly understood, although we might well disagree about the valuation of these various effects. Madaus outlines the influence of standardized achievement testing in the following seven principles:

Principle 1. The power of tests and examinations to affect individuals, institutions, curriculum, or instruction is a perceptual phenomenon. If students, teachers, or administrators believe that the results of an examination are important, it matters very little whether this is really true or false—the effect is produced by what individuals perceive to be the case.

Principle 2. The more any quantitative social indicator is used for social decision making, the more likely it will be to distort and corrupt the social processes it is intended to monitor.

Principle 3. If important decisions are presumed to be related to test results, then teachers will teach to the test.

Principle 4. In every setting where a high-stakes test operates, a tradition of past exams develops, which eventually de facto defines the curriculum.

Principle 5. Teachers pay particular attention to the form of questions on a high-stakes test (for example, short answer, essay, multiple-choice) and adjust their instruction accordingly.

Principle 6. When test results are the sole or even partial arbiter of future educational or life choices, society tends to treat test results as the major goal of schooling rather than as a useful but fallible indicator of achievement.

Principle 7. A high-stakes test transfers control over the curriculum to the agency which sets or controls the exam.

Policymakers, school administrators, and teachers understand these principles well, even if not everyone in these roles would so define the effects of testing.

The purpose of this article is to explore the ethical issues involved in using standardized testing to implement a state-level policy. The context is New York State, which has a long history of using such a strategy, the best example being the Regents examinations. Beginning with grade 3, there is at least one state-mandated test, and often more, during each year of schooling.

---


The specific example in this article is the 4th grade science program evaluation test, which was administered in New York State for the first time in spring 1989. The Bureau of Science in the New York State Education Department (NYSED) implemented a state syllabus in 1985 that outlines the focus of the curriculum. The syllabus is a nonspecific document that does not prescribe activities, textbooks, or teaching strategies. Rather, it generally states goals to be achieved in science education. As we might imagine, a document of this sort is likely to have little effect on what sort of science gets taught. In fact, many teachers are unfamiliar with the state syllabus in science and in most other subjects, as well. The bureau’s frustration with effecting change in science education across the state led eventually to the development of the Elementary Science Program Evaluation Test (ESPET), administered at grade 4, primarily because no state-mandated test was administered at this one grade, although some people also argue that the syllabus has a natural break at this point.

In this context, I discuss ethical issues, but a word of caution is necessary. I believe the people working in the Bureau of Science, and other NYSED personnel, act in ways they believe are in the best interests of schools, teachers, and students. The deleterious effects of their actions are not a product of their evil intentions or neglect. I believe the ethical principles inherent in actions are embedded in the institutional structures within which people operate. Although the Bureau of Science and others in the NYSED should be held accountable for their choices, it is not a simple matter of finger pointing and blame laying. All policies and the strategies for implementing the policies should be exposed to ethical critique, with the goodness and rightness of positions challenged and justified. Ethical critique is particularly necessary for actions taken by public institutions on behalf of the public and in its interests.

THE POLICY OF THE NYSED BUREAU OF SCIENCE

The Bureau of Science’s policy is directed toward accomplishing two goals. to increase the amount of science taught in the elementary schools of New York and to use a hands-on inquiry approach to teaching science. These goals are reflected in the New York State Elementary Science Syllabus in place since 1985. The Bureau of Science has monitored the implementation of the state syllabus, and although some change has occurred in classrooms, NYSED staff believe it is too little and does not respond to the general public’s demands to immediately improve science education.

Any full critique of the ethical issues inherent in using the ESPET would necessarily raise questions about the worthiness of these goals. Are the stated goals and intents of the policy defensible, and on what grounds? For the purposes of this article, I will leave aside this discussion, not because it is unimportant, but because of the consensus on the worthiness of the goals. With few exceptions, people believe more science should be taught to elementary
students and that using a hands-on manipulative approach to teaching is appropriate. Intersubjective agreement alone does not make these goals worthy, but it does make the goals compelling and motivates efforts to implement the policy. We need to raise questions about the worthiness of this policy, if it is unjustifiable, it may misdirect educational effort and resources. For the purposes of the rest of this discussion, I assume that teaching more science in elementary schools using a hands-on approach is worthwhile.

The ESPET was written into the Regents Action Plan, a document that directs New York schools by recommending and mandating certain actions and by clarifying the meaning of the terms in the Commissioner's Regulations. The Regents Action Plan defines the purpose of the ESPET as a means to measure the effectiveness of school programs in the elementary grades and to measure student progress in those skills where evaluation of student performance requires techniques other than written examinations. Emphasis will be placed on evaluating laboratory work, oral expression, critical thinking, and problem-solving skills.

The ESPET was designed to assess content knowledge, inquiry skills, attitudes toward science, and school environmental factors that affect science education. The NYSED mandates only the student-assessment parts of the evaluation, the attitudinal and environmental surveys are optional, to be conducted at the discretion and cost of local districts. The decision to mandate only the student-assessment parts was a compromise made with the NYSED Bureau of Testing personnel who believed the attitudinal and environmental measures were too “soft” and consequently psychometrically and legally indefensible.

The student-assessment parts of the ESPET were designed to explicitly direct teacher time and effort in certain ways. The test consists of two parts: a 45-item objective test (29 items focus on knowledge and 16 on skills) and a 17-item manipulative test. The objective test is a familiar multiple-choice test, but the Bureau of Science is proudest of the manipulative part for its progressiveness and authenticity. The manipulative part consists of five stations with clearly specified tasks intended to assess students’ abilities to measure, to observe and predict, to interpret data, to classify, to generalize, and to make inferences. The form of the test was explicitly chosen to direct teachers’ attention to specific content contained in the state syllabus and to encourage teachers to use a hands-on experimental approach to teaching science.

THE NYSED TELEPHONE GAME

The ESPET is not what the Regents Action Plan says it is. I am reminded of the Telephone Game we played as children at birthday parties. One person whispers a phrase in a neighbor’s ear, and in turn that person whispers in his or her neighbor’s ear, and so on, when the last person repeats the phrase
aloud, it is significantly altered from what the first person said. So too does the ESPET go through several interpretations, coming out sounding much different from what the Regents Action Plan says. In essence, the program evaluation test (which is actually a misnomer because the "test" is a formative process involving several types of data collection from several sources with specified reporting structures that focus on strengths and weaknesses in the curriculum and school environment) has come to be seen as the same as the student assessment given in grades 3 and 6—the Pupil Evaluation Program (PEP). This perception is partly so because the state mandates only the student-assessment parts of the ESPET, all other data collection is optional and done at the discretion of individual school districts. Some districts undoubtedly will implement the whole program evaluation, but most will not. Although the NYSED supplies all the materials needed for student testing, local districts must bear the cost of administering and analyzing the attitudinal and environmental surveys. Although it may not be substantial, the cost is significant for many districts. Also, the simple cost in time, instructional and administrative, works against full implementation. To date, the NYSED does not have a record of the number of districts implementing the total evaluation program, but in attending meetings with the state network of science mentors, I found apparent resistance to doing so, especially from New York City school districts.

If the purpose of the ESPET is truly to evaluate the elementary science program, then testing every 4th grader in New York State every year on exactly the same test items is clearly unnecessary. Some form of matrix sampling could reduce the data collection and analysis burden and would imply an emphasis other than simple student assessment, but the NYSED never seriously considered this option. Reporting ESPET results on New York State's Comprehensive Assessment Report, a sort of report card for the state's school districts, in much the same fashion as PEP results, has reinforced this perception of the ESPET.

Clearly, the message has become garbled—what began as a comprehensive review of the elementary science curriculum has become one more state-mandated student achievement test. So, what is the problem with this reinterpretation? The confusion about the initial message of the ESPET's meaning is more serious than the confusion created in a party game. If the various constituencies in New York State accept a strategy like the ESPET because of its characteristics, and the implementation occurs without those same characteristics, then the districts and schools have been deceived. Even with no malice or intent to deceive, the consequence is still deception. The Regents, the NYSED, the Bureau of Science, the Bureau of Testing, and the schools are all complicitous in this deception. Ironically, the Regents initially

---

intended to hold schools more accountable for student achievement by implementing more PEP tests. Although the state may have made more sophisticated and elaborate plans, schools in New York State ended up with something that looks a lot like more PEP tests.

The similarity between PEP tests and the ESPET is not without consequence because low PEP scores trigger the public to identify a school as inadequate, resulting in significant NYSED intervention in the school’s operations. Although ESPET results do not have this weight (yet), Madaus's first principle reminds us that—right or wrong—the perceptions of school personnel are what is relevant.

Something is wrong when the actual strategy for implementing a policy deviates significantly from the planned strategy and does so in ways that still accomplish the policy's intents. Perhaps this deviation is natural evolution in policy implementation. Perhaps the plan for the ESPET was not sincere. No matter what argument is presented with what rationalizations, it is not right to say that you will be doing one thing and then to do another. Saying you are doing program evaluation when you are doing student assessment is not right.

THE UTILITARIAN VIEWPOINT

The use of state-mandated testing to implement a policy is based on utilitarianism. As a moral theory, utilitarianism defines rightness as maximizing the good, which is typically expressed as happiness, pleasure, or utility. Maximizing total welfare is important, without concern for the individual or the distribution of good within the total society. So, for example, if the greatest total welfare required that a subset of children, let's say physically handicapped children, not be educated in public schools, then a utilitarian view would hold that not educating these children is morally right and even required.

Assuming this ethical view, the situation in New York schools before implementing the ESPET was not perceived to maximize the good. In fact, most people would have seen the situation as undesirable—not enough science was being taught to young children, what was taught was dull and textbook-oriented, and the schools were not producing enough scientists nor a scientifically literate public. To maximize goodness—teaching more science using a hands-on approach (the utility measure)—the state developed and implemented the ESPET. Following this ethical principle, if, in fact, more science was taught using a hands-on approach after implementing the ESPET, then the total welfare of New York State would be maximized.

A survey of New York’s 4th grade teachers indicates that 68 percent agreed the ESPET clarified what was to be taught, 69 percent said they were using more hands-on activities in their science lessons, and about half of the teachers said they were spending more time teaching science than in previous
years. These data suggest that implementing the ESPET has maximized the good for the total society. Thus, if we assume a utilitarian viewpoint, we can accurately say that using the ESPET was a right thing to do to encourage teaching more science using a hands-on approach in elementary schools.

But these were not the only consequences of administering the ESPET: 59 percent of 4th grade teachers thought the ESPET results would be used to evaluate their teaching, 78 percent believed there was much pressure for students to do well on the test, but half the teachers thought the ESPET did not adequately reflect the science curriculum in their schools. Teachers felt pressure to prepare students to do well on the test, and they prepared students by teaching what was on the test in the form it was tested. For example, one station of the manipulative test requires students to test for conductivity using batteries and a light bulb. With the implementation of the ESPET, most 4th graders will have experience with testing conductivity, at least in the way it appears on the test.

The manipulative test has created a financial burden for schools and sometimes even for teachers themselves. For the purpose of testing, schools are expected to provide the equipment according to explicit specifications provided by the NYSED. Many schools must also purchase equipment for instructional purposes. What may seem a minor resource allocation becomes major if the school simply does not have the resources or is not sufficiently committed to elementary science education to allocate its resources in this way. Because 4th grade teachers feel responsible for preparing students for the ESPET, they will personally purchase equipment and materials that conform to those required for the test. The extent of this problem is unknown, but at least in the first year of implementation anecdotal evidence suggests the problem is not isolated. We might imagine that in the future resources will be diverted away from something else to support the demands of the ESPET.

Another consequence of the ESPET is the increased testing burden on both teachers and students. This burden cannot simply be expressed in terms of testing time but also includes time spent preparing students for the test. Observations in elementary schools indicated abrupt changes in what science education was sometime in April, just before the test was administered. This change in curricular focus is part of what Smith has described as the "natural history of the testing event." The burden of testing needs also to be expressed in terms of the performance anxiety experienced by students and teachers. The sort of anxiety associated with high-stakes testing is evident with the ESPET.

---


For the moment, most teachers favorably perceive the curriculum changes that have occurred because of the test. Although they may not like the authoritarianism implicit in this top-down strategy for curricular change, they identify it as the most effective way to get teachers to respond. Because more science is being taught and many teachers are excited by the hands-on approach, the perniciousness of the ESPET is not yet evident. I would speculate that in only a few years, when the test has become the curriculum, teachers will realize that their authority to make curricular decisions has been seriously limited. In the face of much talk about teacher professionalism, these teachers will see their professional judgment eroded by the testing process, a contradiction that Ross more fully describes.¹⁰

The ESPET affects more than just teachers' curricular decision making; it affects their opportunity to identify, for themselves, areas for self-improvement through professional development. The demands of the ESPET imply that all 4th grade teachers should teach science, and in a certain way. Many teachers undoubtedly feel ill-prepared to do so—elementary teachers typically identify science as one of the least important subjects and often do not feel prepared to teach science. Implementing the ESPET suggests that elementary teachers need to be reskilled. The NYSED has decided what specific professional development 4th grade teachers need, although it will provide little institutional support. A strictly utilitarian view, one that the NYSED Bureau of Science essentially assumes, allows us to ignore these consequences, other than the increased time spent on science using a hands-on approach. I contend that we can ignore these consequences only if we consider the quality of life in schools, for teachers and students, relatively unimportant.

OTHER ETHICAL CONCERNS

Utilitarianism is an ethical view embedded in the institutional structures—the various constituencies involved with the ESPET did not agree or decide to adhere to these principles. "Utilitarianism, it is often held, was developed to bring order, consistency, and coherence to our considered moral judgments."¹¹ This ethical imperative coheres with the bureaucratic structures found in most modern institutions. We must recognize, however, that this ethical view is not "natural." Clearly, we can think about ethics and justice in other ways that would entail different actions. I believe that two important related ethical concerns are missing from the NYSED's utilitarian...
rian view, the distribution of consequences among constituents and an ethic of caring.

**The Distribution of Consequences among Constituents**

Even though a utilitarian view might see that the ESPET enhances the total welfare, the benefits vary among the constituents affected by the test and the changes it induces. In particular, I have focused on the consequences of the testing for teachers. We should judge the rightness of any policy-implementation strategy by how benefits (and harms) are distributed—by specifying criteria, other than those implied by utilitarianism, for distributing justice. Shapiro outlines several alternatives—equality, Pareto optimality, majority, minimax, and dominance.\(^2\)

Rawls presents a more elaborate approach to judge the fairness of the distribution of benefits.\(^3\) Essentially, he claims that justice should be mutually advantageous and is possible under the conditions that lead to "justice as fairness":

Justice as fairness begins from the idea that the most appropriate conception of justice for the basic structure of a democratic society is one that its citizens would adopt in a situation that is fair between them and in which they are represented as free and equal moral persons.\(^4\)

We must begin from what Rawls calls the original position in which people are ignorant of their particular circumstances to equalize the bargaining among people. If this hypothetical situation occurs, Rawls suggests two principles will follow:

1. Each person has an equal right to the most extensive system of equal basic liberties compatible with a similar system of liberty for all.
2. Social and economic inequalities should be arranged so that they both benefit the least advantaged, consistent with the just-savings principle, and are attached to offices and positions open to all under conditions of fair equality of opportunity.

Using Rawls's theory, an ethical strategy for implementing NYSED science policy would assume this original position. NYSED personnel, school administrators, teachers, and the public would presumably be equal and shrouded in a "veil of ignorance." Thus, any strategy chosen would reflect basic liberties (e.g., freedom of speech, freedom of person, freedom of property) and would allow social and economic inequalities only if they benefitted the least advan-

---


taged. Would the ESPET have been the choice of strategies for improving science education under these principles? Probably not. The strategy causes teachers anxiety, restructures their work in ways they do not control, and defines science knowledge in limited particular ways. We can identify deleterious effects for students, as well. We might even consider teachers and students to be least advantaged, with limited access to positions of authority, in this situation, in that case, the strategy also violates Rawls's second principle.

An Ethic of Caring

Not unrelated to the concern about distributing consequences is the notion of an ethic of caring:

When this ethical orientation is reflected on and technically elaborated, we find that it is a form of what may be called relational ethics. A relational ethic remains tightly tied to experience because all its deliberations focus on the human beings involved in the situation under consideration and their relations to each other. A relation is here construed as any pairing or connection of individuals characterized by some affective awareness in each. It is an encounter or series of encounters in which the involved parties feel something toward each other. One who is concerned with behaving ethically strives always to preserve or convert a given relation into a caring relation.15

This ethic of caring entails giving our undivided attention to the needs of others. In mature relations, the role of caring and being cared for can be exchanged. This ethic makes sense situationally, not by adhering to any particular notion of rights or duties. Therefore, dialogue and practice are important means we use to act ethically toward one another. The ethic of caring is almost entirely missing from the implementation of the ESPET. Applying this ethic would first specify a particular relation toward students, as well as relations among teachers, between administrators and teachers, between NYSED and school personnel, and among NYSED personnel. Using the ESPET to improve science education ignores these needs almost entirely.

ALTERNATIVE POLICY-IMPLEMENTATION STRATEGIES

I have criticized the use of the ESPET to implement what appear to be worthwhile changes in science education in New York State. I have suggested that this strategy is based on a utilitarian ethical view and that the view ignores how the strategy affects different people in different ways, in uncaring ways. Madaus has enjoined those of us critical of testing to suggest "attractive counterstrategies." For me, an attractive counterstrategy must be conceived with an eye to the process by which the strategy is arrived at and the consequences for all groups involved, not just some overall sense of welfare. So,

I briefly outline a few counterstrategies for improving elementary science education, although their attractiveness may be disputed.

1. The ESPET as originally conceived by the Bureau of Science was a reasonably well formulated plan for program evaluation. Giving schools the opportunity to implement the ESPET as designed and without the burden of publicly reporting student assessment data would encourage formative, locally useful curriculum evaluation. The NYSED would have to relinquish control over the monitoring process or to significantly alter the perceptions of school personnel about the uses of ESPET data.

2. At the least, student assessment should be based on some form of matrix sampling that would provide ample data for curriculum evaluation and be less burdensome to students and teachers while providing a broader base of evaluative information.

3. Elementary science education may need improvement because teachers do not feel comfortable or competent teaching science. If so, then strategies encouraging teacher development would make sense. These strategies, however, must be implemented at the local level, teachers should decide what they need to improve science education. For example, in a particular school, teachers might decide that a specialist should teach all science classes, and they would either hire the specialist or retool existing staff to satisfy the need. Or teachers might decide to infuse science into the rest of the curriculum (analogous to the whole language approach), and therefore all staff members would need to make a particular effort. The strategies used would necessarily vary from situation to situation, but tailoring to particular needs would more likely contribute to serious, productive improvement in science education.

4. Not every elementary school in New York State has a poor science education program. Strategies that consider different schools' needs and desires for assistance in accomplishing the Bureau of Science's goals are cost-saving and cost-effective. The money now used to support massive student testing could be easily diverted to other strategies for improving science education. Schools wanting assistance might be self-identified, or other measures could be employed. The NYSED would be responsible for providing resources without deciding beforehand how individual schools would improve their science programs. The NYSED could provide resources to local school districts in various ways—grants-in-aid, technical assistance, and paid professional development leaves.

CONCLUSION

Something like this ESPET example probably exists in any state in some form. The particulars of the situation may differ, but the ethical concerns remain constant. A utilitarian view of ethics dominants and thus ignores the consequences of testing for different constituents and does not heed fundamental needs to care for one another.
The alternative strategies for implementing a policy do not have the quick-fix appeal of student achievement testing. They are not administratively simple, nor uniform, nor as cost-efficient. But the damage done to schooling, teachers, and students far outweighs these qualities. In the name of fairness and justice, we must begin to truly see teachers as professionals who can and do serve the interests of students.

SANDRA MATHISON is Assistant Professor of Educational Theory and Practice, School of Education, Education 122, State University of New York at Albany, Albany, NY 12222.


The author traces the philosophical roots of the contemporary moral crisis (nihilism) and argues the place in the school's curriculum of both deontological and theological ethics. The first five chapters on value and moral education conclude with an explication of a human rights ethic as a basis for the structure of curriculum. The last five chapters focus on an interpretative, normative theory of curriculum and pedagogy. Throughout the book, the author analyzes, critiques, and synthesizes the work of dozens of leading theorists. His own theory has multiple perspectives on common general education that draws on features from Spencer, Dewey, Peters, Broudy, and Green. Chapter 9, “Knowledge in Education,” proposes a unique seven-strand curriculum of the arts, crafts, and trades; sports and the dance; natural sciences; social sciences; humanities; reading; and mathematics.


The author presents curriculum from a critical perspective and as a contextual social process. This process includes interactions among knowledge, the total atmosphere of the educational setting, and teachers and students. In addition, the author urges that we not view curriculum as an isolated construct apart from its sociocultural and multidimensional contexts of community, state, and national control.

—Therese Ream


This collection of articles on curriculum and other topics related to early childhood teacher education provides a comprehensive overview of content, pedagogy, historical foundations, research base, reforms, professional standards, teacher recruitment and retention, programs outside the United States, current issues, and prospects for the future.