

The Texas Teacher Test

LORRIE A. SHEPARD

AMELIA E. KREITZER

The TECAT, the Texas Examination of Current Administrators and Teachers, is a test of basic literacy that was given to Texas teachers in March, 1986. The test, seen as politically essential to leverage a tax increase and pay raise for teachers, was intended to raise the public esteem of teachers by weeding out incompetents. Teachers expended massive effort in reviewing basic skills and drilling on test format. After two tries, 99% of the 210,000 who took the test had passed. Shop teachers, special education teachers, and coaches were overrepresented among the failures. The costs of district-sponsored workshops and the in-service day to take the test brought its public cost to a sum 10 times greater than policy makers had anticipated. Though most teachers agreed that literacy skills are prerequisite to good teaching, paradoxically, most also reported that being threatened by a low-level test of fundamental skills was demoralizing. Ironically, many think that the TECAT damaged public esteem for teachers because stories about incompetence in teaching and portrayals of teachers' trepidation appeared alongside examples from a very easy test.

Texas is one of three states with legislated mandates to test the competency of practicing educators. In March of 1986, 202,000 teachers and school administrators took the TECAT, the Texas Examination of Current Administrators and Teachers, to see if they could keep their jobs.

The Texas TECAT was selected for focused research because it was a highly visible example of a reform strategy contemplated by other states. Policy makers who wish to improve public education see several options: increasing high school graduation requirements, lengthening the school day, testing students, providing preschool education, changing funding formulae, and so forth. Testing teachers particularly appeals to reform-

minded decision-makers because it is a concrete and decisive action aimed directly at the quality of education in the classroom.

The Texas teacher test was also deemed worthy of systematic investigation because it exemplifies a new genre of tests accompanying the educational reform movement. Unlike large-scale assessments used in the past to monitor the effects of program change, current reformers now use tests themselves as the instruments of change. Tests of this type are potentially powerful interventions intended directly to raise standards either by spurring examinees to improve skills or by removing examinees with inadequate performance. These reforming tests, akin to what Popham (1986) calls high-stakes tests, are not well understood by either measurement specialists or policy makers. Laws to create these tests are passed based on beliefs about their effects; opponents have different beliefs but no more compelling evidence.

This article summarizes a case study (Shepard, Kreitzer, & Graue, 1987) aimed at understanding both the context and the effects of the Texas test. Can lessons from the Texas experience inform policy decisions in other states? What were the educational problems and political context that gave rise to the reform legislation? What did advocates and opponents believe the effects of testing would be? What were the effects of the test? Who failed? What was known about the teaching competence of those who failed? What can be said about the impact of the teacher test on the quality of education and public confidence in schools? How much did the testing program cost and were the benefits worth the cost?

Research Methods

The study involved an amalgam of social science research methods. Key

political figures and informants were identified from newspaper accounts and by asking each respondent for the names of other central participants. Structured interviews were conducted face-to-face with individuals centrally involved in the creation of educational reform legislation and its implementation: The chairman of the House Public Education Committee and bill author, chairman of the Senate Education Committee, Speaker of the House, legislative aides, the Commissioner of Education, presidents and lobbyists for the four teacher organizations, and Texas Education Agency (TEA) directors of assessment. Respondents were asked to describe the political climate and key events that led to the inclusion of a teacher test in the reform legislation; to characterize their own role in the enactment of the legislation; to depict the positions of proponents and opponents; to relate their perceptions of the impact of TECAT; and to give advice to legislators in other states.

Interview sessions were audio recorded, transcribed, and subsequently analyzed in two stages. Transcript segments were coded as answers to preordained questions or as emerging themes or issues. Identifying labels were assigned to new issues to link recurring themes across interviews. For example, "no test, no tax" was a code used to tag references to the political bargain (ultimatum) made in the last days before the passage of the legislation. In later rereadings of the

LORRIE A. SHEPARD is at the Laboratory of Educational Research, Campus Box 249, University of Colorado at Boulder, Boulder, Colorado 80309; AMELIA E. KREITZER is in the Educational Research, Measurement, Evaluation Program of the Department of Education, Boston College, Chestnut Hill, Boston, Massachusetts 01267.

transcripts, subtypes or competing positions were identified. Exemplars were collected of each position and theme.

Stories and editorials from two Texas newspapers were collected systematically. From the *Austin American Statesman*, articles were assembled under the following headings: "The TECAT until April 1986," "The 1984 special legislative session," and "The Select Committee on Public Education." From the archives of the *Amarillo News-Globe*, articles were obtained for three files: "School testing," "1984 special legislative session," and "H. Ross Perot." The same descriptors used in these files guided the collection of a small set of articles from the *New York Times*. Clippings were also obtained unsystematically from seven major Texas newspapers. Prior to the first site visit in May of 1986, newspaper accounts were used to construct a chronology of previous reforms and events leading to the administration of TECAT. Later, a content analysis was undertaken of the two complete newspaper files to determine how public school teachers were portrayed by the press.

Original documents were collected and examined. The Texas Education Agency provided copies of TECAT descriptive materials and the TEA-produced Study Guide, as well as background memos and data. For example, the Commissioner's letter to the State Board was available with the data used to facilitate standard setting. The test contractors had written professional papers describing test development procedures. The Governor's office granted access to the files of the Select Committee on Public Education, which included data such as SAT scores, transcripts of testimony from educator groups, and drafts of findings from the subcommittees. Interview respondents often supplied relevant documents from their files, including transcripts of speeches, instructional materials, and newsletters. If informants referred to data or reports, sources were retrieved. The Public Policy Resources Laboratory, Texas A & M University, which conducts a Texas public opinion poll, was contacted for a complete set of releases on education issues.

A representative sample of 100 Texas teachers and administrators was selected using a two-stage sampling strategy. First, school districts were

stratified by size and a stratified-random sample of 20 districts was selected. Then, educators were randomly selected to represent the stratum proportions. These teachers and administrators were interviewed by telephone, following a structured protocol, during the summer after the TECAT administration. The response rate was 96%. Based on the demographic characteristics of the non-respondents, the reported results could reflect a slightly positive bias. The interviews were recorded, transcribed, and analyzed following the procedures described for the key figure interviews. In the case of the teacher interviews, several of the questions could also be summarized quantitatively, as in the number of hours spent studying for the test.

The authors attended a total of five days of workshop sessions offered by regional Service Centers to help teachers prepare for the test. The sites were selected to achieve geographic and demographic spread. Of the several sites using University of Texas developed video tapes for instruction, only one was chosen. Researchers were denied access to three sites because of the amount of stress involved for the participants who were retaking the test. The preparation sessions were recorded, transcribed, and analyzed. Instructors were interviewed. Workshop participants completed teacher questionnaires on a voluntary basis. Although responses could not be considered representative, the questionnaires provided answers and writing samples from teachers who had failed TECAT the first time.

A representative sample of personnel directors was interviewed in the fall of 1986, by which time teachers were either to have passed TECAT or been fired. In telephone interviews using a standard protocol, directors were asked about the general impact of TECAT, about the teaching qualifications of teachers who failed, and about the incidence of teachers not taking the test. Other questions pertained to the specific treatment of teachers who failed and to the more general handling of teacher shortages in that district. Methods of analysis described previously were again applied to written transcripts of these interviews.

The research methods in this study can be described as both preordinate and responsive (Stake, 1973). Many of

the study populations and research questions were determined beforehand. Other aspects of the study were developed in response to issues and questions identified in the early stages of data collection. For example, once it was known that Education Service Centers played a major role in providing preparation and review for the test, personnel at all 20 centers were contacted and asked standard questions about the instruction provided and the region's population of failed examinees. Copies of instructional materials were requested and schedules were obtained for remaining workshops. Phone calls were made to cable TV stations in several major cities to document the extensive schedules of TECAT broadcasts which had been described in teacher interviews.

The methods of the field work are described in greater detail in the full-length technical report (Shepard et al., 1987).

Political Context

Until 1980, oil had held Texas immune from economic troubles felt by the rest of the country. Texas had gone through the 1970s without a state income tax and without raising the sales tax, all on the profits from oil. But when the world-wide energy glut reduced the price of crude oil, the state's dependency on oil revenues turned the boom to bust overnight. Participants recalled that political rhetoric about correcting the ills of public education was tied to business interests and accelerated when Texas first felt the effects of economic recession.

In 1982 Governor Mark White had been elected after promising to seek a 24% pay increase for teachers, but failed to get the necessary tax increase from the legislature. In 1983 several events came together to create the impetus for major reform with an accompanying tax increase. *A Nation at Risk* and other national reports added to the dismay in Texas about the state of public education. When we asked aides, legislators, and teacher leaders to think back to specific evidence of educational decline, they recalled that: "Texas was near the bottom compared to other states on SAT scores" and "Across the nation the weakest college graduates appeared to be going into teaching."

Also in 1983, failing to obtain a budget increase, the governor ap-

pointed a Select Committee on Public Education chaired by multi-millionaire, H. Ross Perot. Many say it was the power and visibility of Perot that gave this blue-ribbon panel clout not enjoyed by previous advisory committees. He had a penchant for oneliners that kept the Committee's work in the news for a year (e.g., "scheduling academic subjects around band and sports is joke"). He reportedly spent one-half million dollars to effect the work of the Select Committee and hired his own lobbyists to see the reforms through the Special Session of the legislature. Business leaders heard from one of their own that Texas could not hope to compete for high technology investments if northern executives were unwilling to move their families to Texas schools.

The Select Committee heard testimony about the need to upgrade the profession of teaching, to pay higher salaries to attract the best people into teaching, and to keep the most talented from leaving. In the process they collected horrifying stories about the incompetence of some practicing teachers. Apparently one teacher was said to have had difficulty explaining to her class why the weather was so different in Hawaii and in Alaska even though they were right next to each other (in the corner of the map). A formal survey revealed that most educators believed that up to 10% of their colleagues were incompetent. Furthermore, a significant percentage of superintendents, principals, and teachers felt that it was difficult to fire a bad teacher (Sirota & Alper, 1984). Recurring news stories emphasized that a substantial number of teachers lacked basic skills. In Houston, 62% of new teachers failed the Pre-Professional Skills Test; in Dallas, the superintendent explained that he was hiring below standard minority teachers to satisfy a court desegregation ruling, and an equal number of unqualified whites to avoid reverse discrimination (*Dallas Times Herald*, 12/13/83). Informally, committee members shared stories about letters they had received from teachers arguing for pay raises that were peppered with bad grammar and misspelled words.

Teacher testing to weed out incompetent teachers emerged from the Select Committee recommendations as a necessary element in educational reform. A test for practicing teachers

might have been inevitable from that point on, although the advisability of testing was hotly debated in the legislative Special Session. The inevitability of a teacher test was sealed when it became a bargaining chip to leverage a tax increase. Lobbyists and aides heard over and over again the phrase, "no test, no tax," attributed to Stan Schlueter, Chairman of the House Ways and Means Committee. A tax bill would not get through his committee without evidence that higher wages would not go to incompetent teachers. As many legislators voiced it, "We have to be accountable to our constituents. We are not going to pass this kind of tax bill if we can't assure them that there aren't teachers who can't spell or read or write." The accounts we heard of the *quid pro quo* bargain in Texas resembled very closely rhetoric that accompanied advent of the Arkansas teacher test.

The Texas reform legislation was an omnibus bill. It provided for redistribution of resources to poor school districts, higher starting salaries for teachers, a career ladder, statewide textbook adoption, a high school graduation test, the famous "no pass, no play" rule, and a dozen more changes. Among the provisions were two levels of testing for current educators: subject-matter tests and a test of each examinee's ability to read and write. The 1984 Special Session also delivered a 4.8 billion dollar tax increase, the first state increase in 13 years.

The TECAT

Day-to-day implementation decisions further shaped the Texas test. Originally, some key legislators had assumed that many teachers would be exempt from taking a basic skills test because they had already taken standardized tests such as the SAT. ETS withdrew its tests, however, because they had not been validated for the proposed use; nationally known measurement expert, James Popham, warned that even with new validity studies, the state would be vulnerable to lawsuit if teachers did not have equal access to different tests (meeting transcript, 2/1/85). To defend the job relevance of subject-matter tests, many, many tests would be required. "An eighth grade math teacher could not be given the same test as a teacher of calculus." Although the Texas Education Agency was in the process of developing ap-

proximately 30 subject-matter tests for new teachers seeking certification, legal counsel advised the State Board of Education that these tests would not be defensible for practicing teachers. The board reported to the legislature that 15 million dollars would be required to implement the subject-matter tests. Commissioner Kirby estimated that a basic skills test alone would weed out 80 or 90% of incompetent teachers (*American Statesman*, 2/2/85). The state board proposed and the legislature accepted a basic communications test as a good faith implementation of the testing requirement; 6.5 million dollars were appropriated to develop this test and appraisal procedures for the new career ladder.

A new test, The Texas Examination of Current Administrators and Teachers, was developed to assess the minimum reading and writing skills "that practicing educators need to perform adequately in their jobs." The reading test was comprised of 55 multiple-choice items that measured recognition of details and comprehension of the main ideas, job-related vocabulary, ability to distinguish fact from opinion, reference usage, and inference. The writing test included both a short composition (150 words) and a multiple-choice portion. If examinees unambiguously passed or failed on the composition, their multiple-choice answers would not be considered. Examinees who turned in a marginal essay, however, had to pass the 30-item multiple-choice portion covering mechanics, sentence formation, and English usage in order to pass the writing test. School personnel who failed the reading or writing part of the test in March of 1986 would have one chance in June to retake the portion they had failed. Subsequent retakes were permitted, but not in time to forestall being without a certificate in September 1986. (A sample failing essay is shown in Figure 1.)

TEA staff and their contractors devoted extensive effort to constructing a test that could withstand political and legal scrutiny. A statewide Advisory Committee and a Bias Review Committee were established. A statewide job-relatedness survey of 4,000 educators determined which skills were included and 1,000 educators reviewed potential test items for appropriateness and importance. The commissioner proposed that passing standards be set at 75% after reviewing field-test passing

rates for each score and recommended standards from the advisory committee and several survey groups. The commissioner's standard would have failed 12% of the field test group; however, staff estimated that the actual failure rate would more likely be 5% because teachers would study for the real test. The board adopted the commissioner's passing rule in January, two months before the March testing.

Massive Preparation

One of the most unexpected findings from our research was the monumental effort that went into preparing for the TECAT. As soon as the test specifications were available, the Continuing Education Division of the University of Texas at Austin, in cooperation with the Texas Classroom Teachers Association, developed a review course and a 300-page study book. They trained 130 presenters who in turn instructed 89,000 teachers in one- and two-day workshops. According to Dr. Shirley H. Crook, project director, "it was a major undertaking, logistically." On some weekends 50 workshops were going on, with 20-400 teachers in attendance at each. In addition, the University of Texas, in conjunction with the Austin Independent School District, developed 12 video tapes covering TECAT skills. These were distributed throughout the state. Most of the 20 regional service centers and many school districts purchased the tapes and checked them out to teachers or used them as the basis for group review sessions. One superintendent kept a VCR in his home set up for teachers "night and day" while his wife served popcorn. The videos were acquired by public access TV stations in major cities and shown repeatedly before both the first and second TECAT administrations. For example, in Austin the tapes covering different skills were shown every half hour for 12 hours a day for 30 days preceding each testing date.

All four major teachers organizations developed materials or conducted workshops to ensure the success of their members. (One union forbade the use of their materials by nonmembers; other were more generous.) The largest group, the Texas State Teachers Association, estimated that 65,000 teachers attended their workshops. Nearly every school district in the state provided test preparation oppor-

tunities. In some cases, they used the Study Guide developed by the Texas Education Agency and hired English teachers to conduct inservice sessions. Some districts made arrangements with local colleges and universities for review classes. Many districts, including large districts such as Houston, Corpus Christi, and San Antonio, paid the workshop fees for their teachers. The 20 regional service centers likewise committed themselves heavily to preparing educators for the TECAT. Several developed their own materials and practice tests; some arranged traveling workshops to reach remote areas. Many regional centers hosted six-hour video programs delivered by satellite through the Texas Interactive Instructional Net-

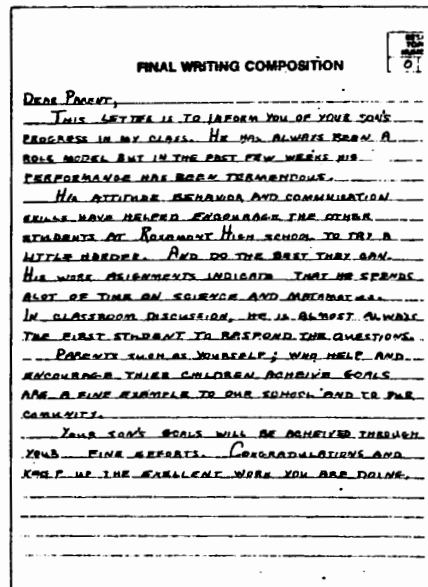
mood that followed the announcement, the Commissioner of Education expressed his delight, "The best news of all is that the TECAT scores are in and teachers did fantastic. Their performance on this basic skills test actually surpassed our expectations." Indeed, teachers had done better than would have been predicted from the field test data, where 12% had failed. Staff at the Texas Education Agency believed that all the training had had an impact, accounting for the very high passing rate on the real test. Many educators and politicians whom we interviewed about the TECAT believed that it forced teachers to learn basic skills essential for proper functioning in the classroom.

In our analysis of transcripts from preparation sessions we classified instructional talk as content teaching, legitimate teaching to the test, and questionable teaching to the test. Content teaching included overviews of test content, substantive presentations of rules of grammar, principles of good writing, and detailed explanations or examples. All practice time was classified as content teaching, whether on sample sentences or specific test items, unless a test strategy was being emphasized rather than a substantive rule. Our conception of legitimate teaching to the test follows the Test Standards of the American Psychological Association, which urge that test takers be informed of any strategies that are "unrelated to the construct" but "influence test performance" (APA, 1985, p. 27). Many of the topics covered in the University of Texas course and materials fall into this category, for example, familiarization with test format, scoring rules, advice about guessing strategies, and anxiety reduction techniques. Examples of this type of teaching to the test are given by these excerpts from workshop presentations:

- You don't want to leave any empty spaces. There are no penalties for guessing on this test.
- The TECAT will cover only two uses of the semicolon, both of which involve compound sentences.
- Remember you only have to know it's wrong. You don't have to know why, you don't have to correct it. Just know that it's wrong.

Each of the writing workshops we attended stressed being concerned with the appearance of the essay (appro-

FIGURE 1
Example of a Failing Composition
 (From the TECAT Passing Standards, TEA, January, 1986)



work (TI-IN). One director commented that there could not have been a single teacher in the state who wanted formal review but could not find it.

Basic Skills and "Teaching to the Test"

On the day the first TECAT results were announced, and an impressive 96.7% of the teachers passed, the test author, Dr. James Popham of IOX Assessment Associates, congratulated Shirley Crook for the incredible impact of the University of Texas' instructional program. Sharing in the euphoric

priate length, legible handwriting, and correct letter format) so as to make a good impression on the scorer. Although this type of preparation should not be considered unfair or inappropriate, it would be hard to argue that the substantial time spent in these activities was really teaching teachers essential basic skills. In fact, the great majority of teachers interviewed in our probability sample said that workshops had helped them by familiarizing them with the test format and sample content rather than teaching them skills. Only three of the 96 respondents said they learned something new, specifically, rules regarding punctuation and the use of pronouns.

At some point legitimate teaching to the test crossed over an ill-defined line and became inappropriate. For example, after explaining that the writing samples would each be graded in one or two minutes, one instructor explained that "it's better to paragraph in the wrong place than not to paragraph at all" (because at least it would look right to the grader). This and other examples of questionable teaching to the test went beyond helping the examinee "show what he knows." Instead the strategies had the effect of helping the examinee "hide his ignorance" or use the multiple-choice format to "pretend to know." Although these strategies are not illegal (they are permitted by the test), their use clearly distorts what the test can claim to have measured.

The most widespread example of this second type of teaching to the test involved exploitation of the test specifications, published by the Texas Education Agency, to "psych out" the multiple-choice test questions. The TEA Study Guide explained how the wrong alternatives would be constructed for every type of question. For inference questions on the reading test incorrect answers would be of the following types:

- **Inaccurate**— A statement that is contradicted by information in the reading sections;
- **Unsupported**— A statement that may sound reasonable, but does not necessarily follow from information in the reading sections;
- **Irrelevant**— A statement that is in no way logically true based on information in the reading selection. This inference often introduces information not included in the section.

In sessions we attended, teachers practiced identifying irrelevant and speci-

fically contradicted answers, so as to arrive at a correct choice among alternatives by a process of elimination. We came away thinking that teachers who were really struggling with inference would now be able to pass the items but would be unable still to recognize two valid inferences from the passage.

Similar strategies for ruling out wrong answers were encouraged for main idea, detail, fact and opinion, and even vocabulary items. The University of Texas tapes included the following information from the test author: *... something very special will occur in the answer options for the FACTS and OPINIONS questions. You'll want to listen carefully to the following information because it virtually ensures success on this TECAT section. Of the four answer options, two will be fact and two will be opinion always. Of the two fact statements, one will appear in the passage and one will not appear in the passage. Of the two opinion statements, one will appear in the passage and one will not appear in the passage. This information has some very important implications First of all, if you are asked to identify a fact, the first thing you can do is simply ignore the two opinion statements. With the two remaining facts you merely have to determine which one of those two happen to appear in the passage. The same would work for the opinion statement. . . .*

Of course the examinee still had to do the basic sorting, still had to recognize fact and opinion. But if a teacher were having trouble making the necessary distinction, wouldn't it make it easier to know that there were always two of each?

The extent of teaching to the test varied greatly. The video-taped presentation had the greatest proportion of content teaching; in the worst case, content was taught less than half of the workshop day. Although content was emphasized in most of the workshops, the widespread availability of "test-taking tricks" has to be considered as a partial explanation for the extremely high passing rates.

TECAT Passing Rates

The data from the first administration of the TECAT are presented in Table 1. The passing rate was 96.7% statewide but with a disproportionately higher failure rate among minorities, especially black teachers and administrators. There was also a pronounced age trend in the data; older teachers

TABLE 1
Performance on the TECAT:
March 1986 Administration

(Individuals who did not report their ethnic group are counted in the total but are not included in the separate categories.)

BY ETHNIC GROUP		
	Number Tested	Number Passing (Percent)
Hispanics	24,685	23,195 (94.0%)
Blacks	15,681	12,802 (81.6%)
White/Other	156,505	154,838 (98.9%)
TOTAL	202,084	195,505 (96.7%)

BY AGE GROUP		
Age	Number Tested	Percent Passing
0-29	38,971	97.8%
30-39	74,706	97.6%
40-49	52,349	97.2%
50-59	29,985	94.0%
60-65	4,615	90.5%
Over 65	721	87.7%
Not Given	384	87.2%

failed at higher rates.

Data reported by institutions granting college degrees showed considerable variability. For example, graduates from the University of Texas at Austin, at San Antonio, and at Arlington passed the TECAT at rates exceeding 99%, whereas numerous colleges in Texas (mostly small private institutions) had passing rates as low as 55%. Graduates of out-of-state colleges had a 97.9% passing rate. Early in our study, one union spokesman suggested that it was this extreme variability in the quality of graduates from different institutions, especially traditionally black colleges, that had been an underlying impetus for the TECAT.

The second administration of the TECAT raised the final passing rate to 99%. Of the 6,579 teachers who failed the test in March, 4,704 retook an equivalent version of the test in June and passed; only 1,199 teachers failed a second time; 676 teachers did not sign up to be retested. We heard vague talk about teachers who decided to retire early rather than face the pressure of the test; but the number of retirees was not appreciably higher in the districts surveyed.

In an effort to understand what kinds of teachers had been weeded out by the test, we relied first on accounts from personnel directors and the representative sample of teachers. We were seeking corroborating evidence that the deficiencies of the failed individuals had been known in the school or district. Three points support the argument that the test got rid of incompetent teachers: nearly all teachers indicated that the TECAT was a fair test of literacy skills essential for good communication in the classroom; teachers "admitted" that the one or two failures they knew personally used poor grammar in their day-to-day conversation; and personnel directors classified the teachers who had been fired after failing twice, as "average" teachers—they were neither exemplary nor very bad. Only rarely (accounting for less than 5% of the failed teachers known to the respondents), did we hear of a woefully incompetent teacher who should have been fired years ago; "the test finally got him."

On the negative side, some teachers were fired because of the test who should not have been. We examined data for districts where the passing rate was 85% or less on the first testing. These districts fell into one of three categories: districts comprised of group homes for the mentally retarded or emotionally disturbed; heavily impacted minority districts (3 districts); or small rural districts with fewer than 30 teachers altogether. Teachers of mentally retarded children were also identified by several personnel directors as those they had lost because of TECAT. As a group, vocational education teachers had a disproportionately high failure rate. Several personnel directors noted that the loss of these individuals seemed particularly unfair because, to be certified, they had never been required to be college graduates. Often teachers who were interviewed expressed regret over a shop teacher who had been fired, "I know he doesn't speak proper English, but he really knows machinery; and he's so good with the kids." Similar ambivalence regarding the legitimacy of the test versus the value of a colleague was expressed about many physical education teachers and coaches, about bilingual education teachers, and a few kindergarten teachers. As part of the cost analysis discussed later, TEA supplied

data which corroborated the overrepresentation of special education teachers, vocational education teachers, and coaches among the failures.

Teacher Morale and Public Confidence

Before our first visit to Texas, we had read newspaper accounts about teacher protests and about how insulted teachers felt at having to take a literacy test. Early on, leaders of teachers' organizations told us extensive stories about the anxiety and disruption the test had caused. But we had also been told by TEA staff and legislative aides that these stories were mostly union "hype." Legislators who had sponsored the reform legislation believed that militant union leaders did not speak for the majority of teachers. They knew teachers personally back home who were quite willing to take the test, if it served to get rid of the few incompetents who were giving the profession a bad name.

Interviews with scientific samples of teachers and personnel directors were intended to give a more representative picture. But even without the filter of politics or media sensationalism, we were told consistently that the test had created tremendous stress and bitterness. Most compellingly, the 20-25% of teachers who did not themselves feel threatened by the test nonetheless described its negative impact on the majority of their colleagues. Simultaneously, the majority of personnel directors said that the TECAT had had no real effect in their district because virtually everyone passed, but it had generated negative attitudes and made teachers feel degraded. "We had a very bad year."

Teacher interviews resulted in more than 1,000 pages of transcripts; more than 100 pages were in response to these questions:

Did preparing for the TECAT make you a better teacher? Did preparing for the TECAT have any negative effects on your teaching during the past year? (And later in the interview:) What has been the effect of the testing program on teachers?

Only 5% of educators said that preparing for the TECAT had made them better teachers or administrators. Half said that studying or worrying about the TECAT had hurt their teaching because of the time it took, because of the stress they were under, or because they were less willing to give time to extracurricular activities.

The following verbatim quotations typify the reported effect of the testing program.

- *I think mostly what I found negative was the way I saw some of my peers and also some of my superiors, those who I looked up to and respect, older people, become concerned and a little bit worried about the thought of having to take a test to enable you to secure your position.*

- *... some of my fellow teachers just went into orbit about the test, they were so anxious about it. I think for no reason, in some cases. Then it had a negative effect on their teaching, and, as a result, I watched their frustration build, and I began to wonder what the test was about.*

- *The morale really dropped. I have never heard so many teachers say, "If I could find another job." "It's about time for me to retire." Things like that. It's really hurt.*

- *I don't know how else to say it except it was just a humiliating experience. One of the worst experiences I may have ever been through as a teacher.... I begrudge the time that I had to take to study for it, to worry about it.*

- *We usually get really enthusiastic about different units that we're teaching, things we're going to study and things like that, and we just couldn't get ourselves up for it. We felt, you know, that people thought we were incompetent, and the kids, even in first grade, they would comment.*

- *Everybody felt like [we] were incompetent—from the kids to the governor. It was deflating to us. And everybody just felt really, really down. I think the governor's idea was that he was going to prove us competent, so teachers would go for him. Well, it backfired on him.*

- *I think it has given a kind of a bitter attitude. And, I think it's been kind of just an embarrassment. You know, they've had sample questions on the t.v. and they're so simplistic that it's almost a joke.*

- *Low morale. It just really socked them in the stomach. Many people were very nervous, uptight, concerned about it, especially the ones who had taught for over 20 years. "What will happen if all of a sudden I'm found unsuited, unfit, without credibility?..."*

The above responses represent over 90% of the teachers interviewed. Others said there was no effect on teachers or gave a positive reply such as the following:

I think that there definitely were some people, I include myself, who were pressed to learn a little bit more about the language; and that's a good thing. There were a lot of bad things, especially for those who pro-

bably knew just about everything they had to know about the language. I'm pretty sure that everybody went through a very stressful time.

To gain some perspective on the negative feelings expressed, we should note that nearly 70% of teachers said that the reading and writing skills measured by the TECAT were prerequisite to being a good teacher. "If you can't do these things you shouldn't be in the classroom." Thus, there was a large discrepancy between feelings about the underlying principle, that all teachers should be literate, and the test itself. But, to be required to take a reading and writing test when they already held college degrees made teachers feel less, not more, professional. To make matters worse, students and letters to the editor persisted in calling the TECAT a competency test, when even the governor admitted it was only a literacy test. Humiliation and embarrassment occurred because media publicity invariably portrayed teacher protests alongside examples that made the test seem laughably easy. High anxiety was created because so much was at stake; many, many teachers said that feelings would have been different if failing the test meant taking a college refresher course rather than losing their job. The pervasiveness of these themes in the two representative samples led us to conclude that TECAT consumed the attention of educators in Texas for the 1985-1986 school year and that it had a devastating effect on teacher morale.

Perhaps a more serious effect of the TECAT was its harm for public opinion. Many teachers felt that the test and accompanying publicity actually worsened public confidence in education. In fact, teachers were about equally divided on this issue. Approximately half of the teachers interviewed said that the test had done what legislators had intended, it had gotten rid of the bad teachers and proven that the majority are competent. The other 50% of teachers felt that all teachers had been made to seem less competent and that the 99% pass rate made the whole thing a joke.

The following sets of quotations characterize the two conflicting positions:

Teacher Position One: The TECAT proved to the public that teachers are competent:

• Now that's where I think something good has happened. I think the public realized maybe a little bit more how hard teachers work and really that they're bright people.

• I hope that it would show the public that we, the majority of teachers in Texas, are not illiterate.

• Well, it may have been a star in our crown. Teachers probably take the brunt of society's ills. And the fact that 98% did pass may have impressed some people who are always complaining that their kid's teacher's probably the stupidest person on the face of the earth.

• I think that the public was surprised to find so many teachers could pass the test. Perhaps they think a little higher of teachers.

• In the school district that I'm teaching in, most of the parents and the public had confidence in us to begin with.

• I think it's important that we let the public know that we are good teachers and that we are teaching them something. If the children are dropping out or they're not learning anything, it's not our fault, it's their fault.

Teacher Position Two: The TECAT was a joke. Now the public has a lower opinion of teachers than before.

• It was ridiculous. They think this was all a farce.

• I think they were hopeful that this was going to weed out incompetents. Every student, including myself, has had a bad teacher. And I think they were hoping in one broad sweep they could eliminate those who were not as professional as they should be. And I think they were disillusioned because of the publicity afterwards—the way the press chose to characterize the type of test it was and how easy the test was.

• I think it's negative. The teachers were behind the 8-ball. If they didn't do well on the test, then obviously they weren't good teachers. And if they did well on the test then the legislature—the first thing they said was, "Whoa! This test was too easy; we passed too many." And, therefore, you were dead if you did and dead if you didn't.

• The public does not like the classroom teacher. People feel that we were mealy-mouthed, that it was wrong for us to be angry over a test. They feel we always whine, wanting more money.

• Some of the jokes circulating now consisted of several pages of ridiculous, very ridiculous things. Say, for instance, "find your way through a maze and the way is outlined very dark." It was trying to emphasize how dumb the questions were, but

I don't think they were that dumb. . . . You know, it was funny for a joke but it's not funny for teachers, it really demeans the teachers.

• I don't think it's had any effect whatsoever on the general public. I mean, half of them are saying, "Gee, that was ridiculous. It was a waste of money. I knew they'd do good." The other half is saying, "Gee, that was a waste of money, the test was too easy so I knew they'd all pass."

Trying to assess what the public really thinks is difficult. Even with representative survey data the results are sometimes internally inconsistent. In 1985, 54% of Texans gave their schools a grade of A or B, a much higher percentage than in the Gallup Poll nationally. In the same survey, however, 79% said they were in favor of competency testing for teachers (The Texas Poll, 1986). The tone of most newspaper articles was consistent with the more pessimistic account. In the commentary, we will discuss how "both sides" contributed to negative publicity that may have harmed the public esteem of teachers.

Cost Analysis

How much did it cost to test every teacher and administrator in Texas? The contracted cost was \$4,833,000 in special funds to develop, administer, and score the TECAT. In addition, the Texas Education Agency subsidized the appropriation by assigning regular assessment staff to the project for an estimated cost of \$232,500.

A summary of the public costs of TECAT is shown in Table 2. The largest cost was the use of a teacher inservice day to take the test. We have also approximated the publicly sponsored costs of providing preparation for the test.

The total tax-supported cost of TECAT was \$35.5 million. The analysis summarized in Table 2 should be considered conservative in that costs were only included if there were data to support the estimate. Additional costs very likely occurred without our being able to represent them in the analysis.

Realistic cost data are important, however crudely estimated, because actual costs were an order of magnitude greater than the anticipated cost of testing. TECAT was expected to cost about \$3 million. This was the number found in the cost estimates of the Select Committee and was still the

figure used when the State Board discussed the feasibility of implementing the legislation by testing every teacher with a Texas-developed test. In most cases Select Committee staff and Comptroller's staff who computed cost estimates for proposed reforms included increases for both state and local jurisdictions. This was not done, however, in the case of teacher testing; therefore costs such as closing school on TECAT Monday were not included. A one-time test for practicing teachers was considered to be one of the cheapest of all the likely reforms. Data on the real public cost of TECAT indicate that it was an expense more on the order of a programmatic intervention such as a proposed pre-kindergarten for disadvantaged four-year olds, rather than an inexpensive item fitting within the error of the estimates for major reforms.

In our original report we also included estimates of more than \$42 million in private costs, including "wages" for teacher study time (at an average of 12 hours), and workshops, materials, and score reports purchased by teachers. The private cost analysis was controversial, not because the figures were inaccurate but because it was considered inappropriate to add personal expenses (especially study time, which had not required an outlay of cash) to an analysis of public policy decisions. In evaluation research, however, it is desirable to assign dollar values to

hidden investments of time and resources as a way of representing the opportunity cost of policy decisions. Although these are not tax dollars, various state policies have the capacity to command that teachers contribute personal resources to the public arena. For example, all teachers could be required to take a professional development course at their own expense; or each teacher might be required to tutor an individual child for 20 hours. The private costs of TECAT were large. Its benefits should be judged in comparison to alternative uses of the same resource.

When TEA assessment staff were invited to review our preliminary findings, they found that our estimates of the public cost of TECAT were credible. The Commissioner and TEA staff noted, however, that our analysis had given no credit for the public funds saved by firing incompetent teachers. To accomplish this analysis, TEA provided detailed information about the job assignments of the 1,950 educators who were removed from the system by TECAT; these individuals either failed twice or left education after failing once. Because of a dispute over the uniform validity of prerequisite skills (see Commentary), we were unwilling to count as successes of the program removal of vocational educators, special education teachers, staff at group homes, P.E. teachers, or kindergarten teachers. The data confirmed that these groups were overrepresented among the failures. But TEA staff argued compellingly that many failures held mainline teacher jobs and were directly responsible for the academic preparation of students. Given the low level of their own skills, it is hard to believe that they could do a good job in teaching basic skills to students. For example, 383 failures were regular elementary school teachers; 22 failures were even secondary school English teachers.

The individuals who had been removed by TECAT were divided into a total of 887 academic jobs versus 1,063 nonacademic jobs. In addition to physical education and industrial arts, we classified music, art, ESL, health teachers, and school counselors as nonacademic assignments. Principals and superintendents were counted as academic assignments along with all regular elementary teachers and secondary teachers of academic subjects.

The 887 firings represent success in the intent to remove incompetent teachers from the classroom. The average salary paid to all certificated personnel in Texas in 1985-86 was \$23,765. If this amount is incremented by 20% to allow for benefits, then it could be said that the annual cost of these 887 incompetent teachers is over \$25 million. As a result of TECAT, this amount of taxpayer dollars will no longer be wasted, and hence is a savings which compensates for the public expenditure on the test. Furthermore, it can be argued that the firings represent recurring savings (less so as retirements and normal attrition would occur), but TECAT was a one-time expense.

Commentary

The positive case. The affirmative side for recertification of incumbent teachers has been presented at a meeting of the American Educational Research Association by test author, Dr. James Popham, and by Texas Commissioner of Education, William Kirby. Popham (1987) argued that society has a right to certify the competence of professionals, especially when children, as clients, have no option but to accept the services of the education professional assigned to them. The stakes are high. "A youngster assigned to a poor teacher as a first grader may become a poor reader, a decisively worse reader at second grade, and perhaps a life-long low achiever. . . . Every child in America's public schools has the right to be taught by a literate teacher." Teachers who cannot read and write adequately themselves serve as terrible role models. They are not likely to require written assignments from students; they cannot comprehend student work; nor can they upgrade their own knowledge by reading professional literature. "Being able to read and write is clearly not the same thing as being able to teach, however, it is in our view a necessary precursor."

Commissioner Kirby (1987) described the need for educational reform; Texas was at the "bottom of the barrel in the nation at risk." Six hundred million dollars in increased teacher salaries were tied to the test which would assure the public "that teachers in the classroom could at least read and write." Because of the test there are 10,000 teachers no longer in Texas classrooms who were there before (approximately 2,000 who failed twice or

TABLE 2
Summary of TECAT Public
Cost Analysis

Test development and administration:	
Nominal cost	\$ 4,833,000
TEA Staff	232,500
Teachers' inservice day to take the test (202,000 teachers at \$130/day)	26,260,000
Local school cost in supplying test sites (1108 site days at \$125/site)	138,500
Preparation workshops and review: Costs to districts and Education Service Centers	
Inservice development or district paid-for workshops. (100,000 teachers × \$30)	3,000,000
Information services and staff time (210,000 teachers × \$5)	1,050,000
Sites for workshops (800 site days × \$125/site)	100,000
TOTAL PUBLIC COST	\$35,614,000

failed once and did not retake, and 8,000 who never showed up to take the test). Kirby asserted that the public attitude toward education has improved and legislators are more willing to spend money on education, even in the face of a \$3 billion deficit.

The negative case. The case against testing was argued most loudly by teacher organizations. "You can't measure competency with a paper-and-pencil test," they said. Some legislators concurred that a test for current teachers was ill advised. Senator Carl Parker, Chair of the Senate Education Committee, called the testing legislation a full employment bill for lawyers:

Somebody tell me how you can draft a test to tell you how to be a good Senator. You can't do it. Somebody tell me how you can devise a test to tell you whether or not someone is an able kindergarten teacher. It can't be done. . . . We're going to give a test to see if they can add two and two and four and four; and if you can do that well, then you can go back to teaching Calculus. We will have insulted every math teacher in the state and we will not have improved the quality one iota.

Test validity, passing, and failing. The argument for and against a literacy test for recertification of current teachers is really a dispute about validity. Is there a level of reading and writing proficiency that is so fundamental to all teaching that without it a teacher could not be competent? Logically, yes, though the level is hard to specify. Must the level of the teacher's skill be sufficient to read student work? Certainly. Must each teacher comprehend at the level of professional literature? The answer is less certain. The higher up the continuum of language skills one goes, the harder it is to say that the skill is essential to good teaching. All pass-fail tests have the problem of where to set the cutoff; the dilemma is much debated in the measurement literature because individuals immediately on either side of the standard are essentially indistinguishable from each other. But, proposing a literacy test to measure teaching competence greatly exacerbates the normal standard setting quandary, because one construct (literacy) is substituted for the other (teaching competence). All parties agree that the substitution works only at the low end; i.e., the literacy

demands must represent the *minimum* for competent teaching.

Logically, then, there is a validity ceiling that creates a downward pressure on the substance of the test. Content is included in the test only if it receives an endorsement by a high percentage of educators (Yalow, 1986). To ensure due process, abundant opportunities to study are provided at public expense and detailed specification of the test content are given out, down to the format for wrong answers. Resisting this trend, the Commissioner urged the Board to set passing standards higher than recommended by various education groups, to anticipate the effects of intensive remediation. But the Commissioner and his staff only projected that the failure rate, after studying, would decline from 12% to 5% by the second testing. Instead, the failure rate was less than 4% at the first testing and less than 1% by the second try. Opportunities to retake the test also work to degrade the implicit standard of the test, because the rules are asymmetrical. Individuals who just barely pass because of measurement error or having learned the tricks of the test are not subject to reexamination. We do not disapprove of these procedures; they are essential to fairness. But policy makers should understand these basic features of a minimalist test. If an individual will be denied his livelihood on the basis of a test, errors must be fairly egregious before they merit flunking. Large numbers of individuals with marginal literacy skills will pass.

Because TECAT necessarily had to be a lowest common denominator test, the results were poorly matched to original political intentions. Politicians had expected the failure rates to be on order of 10,000 after remediation, not 1,200¹. Legislators had wanted to weed out social studies teachers who were deficient in American history and elementary teachers who didn't know the location of Alaska. The TECAT standard was too low to touch these teachers. If TECAT substance had been harder, or the standard higher, its validity could not have been defended.

At the same time that some incompetent social studies and English teachers surely passed, some teachers with badly needed skills were removed. More than half of those eliminated by TECAT were in vocational education, special education, P.E., kindergarten, health, and counseling. They were also dispro-

portionately minority teachers in districts with high concentrations of minority children. These failures, which were often lamented by personnel directors, are merely the other side of the validity-standard setting dilemma. We refer to it as the *fallacy of uniform validity*; one test and one standard simply cannot be equally relevant for all jobs. We do not wish to be apologists for failed teachers. However, we would be willing personally to have our children taught by shop teachers who failed the test by three points, but dismayed to find them instructed by a history teacher who passed by three points.

Overreaction and demoralization. When Ross Perot's Select Committee first proposed a test to eliminate incompetent teachers, the most frequent number used to estimate the extent of incompetence was 10%. Thus, had they been able to identify themselves, 90% of teachers should have felt relatively safe. They might have spent one or two hours studying and taken the test without much notice. Instead, there was intense studying and no other talk but TECAT for six months.

Why wasn't the teacher response more in proportion to the real risks? First, the public talk surrounding TECAT was tough; secondly, there was a great deal of misinformation which fostered the impression that the test might be unfairly hard and esoteric. Softer legislation, which would have made a test contingent on the appraisal process and provided for remediation, had been rejected. Teachers and administrators had two tries to pass the test and then would be out of a job. One school district wished to fire all teachers who failed once but was prevented by a court order. Because of the short time line for test development, official study guides were not available until the fall of 1985. Districts were in a hurry to start studying and made up their own tests, which were not necessarily at the appropriate level. For example, the *American Statesman* (1/20/85) reported, "Mock competency test flunked by half." At the same time, the first administrations of the PPST to prospective teacher candidates produced failure rates (approximately 40%) that would have been alarming to practicing teachers. Thus, instead of only 10 or 20% worrying about the test, only 20% of our sample felt immune

from worry. The humiliation arose because a literacy test, not a professionally uplifting exam, was the source of the anxiety. To be immersed in studying for a basic skills test that would determine their future had a demoralizing effect, even though the perceived threat was out of keeping with the intention of the test.

Negative portrayal of teachers. Prompted by the disagreement among educators about the effect of TECAT on public opinion, a systematic content analysis was conducted of two complete newspaper files to examine the portrayal of teachers.

Teacher incompetence in Texas had been a recurring education story culminating in earlier 1981 legislation to test entering and exiting teacher candidates. It was a dominant theme during Select Committee hearings. Perot made headlines with the charge that, "The Dumbest People in College Study to be Teachers" (*Amarillo Globe*, 5/17/84).

A second theme in newspaper stories had to do with the unprofessionalism of teachers. They were portrayed as self-interested and anti-reform. Perot called education lobbyists " 'pick-pockets' who want taxpayers to 'send more money but skip reform' " (*American Statesman*, 4/17/84). Later in the special session, teacher groups were cast as mercenaries when they lobbied against equalization. By protesting the test on the basis of their life-time contractual rights, teachers' unions appeared to defend incompetence. After having agreed to the reform package, three of the teacher groups withdrew in a huff, which legislators denounced as capricious and unethical.

In the year leading to TECAT, teachers were in the news filing lawsuits, studying for the test, picketing the governor, and complaining about score report labels and proctor qualifications. Interviewed after the test, individual teachers told reporters that it was insultingly easy. In stories that followed, legislators complained that the test had been too easy and the passing rate too high.

We cannot assess what individual citizens believe about the TECAT. We can say that for a period of three years, news stories presented an undignified and unprofessional picture of teachers.

Ironically, one of the intentions of the reformers had been to raise the

esteem of the teaching profession. The Select Committee had been dismayed by a survey of 1983 Texas honor graduates; only 12% said they would consider teaching in public education. The status of teaching was to be raised by increasing salaries and by removing incompetents. Yet because of the punitive tone of the test and the behavior of teacher organizations, young students might now see teaching as a less admirable profession than before.

Cost-benefit. The public cost of TECAT, counting taxpayer supported review sessions, was over \$35 million. The real cost was roughly 10 times greater than original estimates, which pictured teacher testing as a cheap reform.

The direct effect of TECAT was to eliminate approximately 2,000 of Texas's 210,000 teachers. TECAT also had the effect of drawing approximately 180,000 teachers into review of rules of grammar. By all accounts, however, 90% of teachers were not in need of such review.

Unanticipated side effects of the testing program were the negative impact on teacher morale and, potentially, the harm done to the public image of teaching by negative publicity.

The public and private costs of TECAT represent opportunity costs that should be weighed against other ways of achieving the same ends or expending the same resource. For example, how might the investment in professional development have been spent, if not on TECAT? What other policies might have fostered the removal of 1% of the teaching force judged to be incompetent? What if the state had set up a fund to support the administrative and legal costs of dismissing bad teachers? (For an account of the normal procedures and impediments to removing incompetents, one of which is feared legal costs, see Bridges, 1986).

Postscript. When the study began in the spring of 1986, legislation to test practicing teachers was pending in two states and talked about in others. Today, these actions are neither passed nor pending. We sense, as does Dr. Chris Piphon who monitors these things for the Education Commission of the States, that there is less enthusiasm to jump into teacher testing now than 18 months ago. We can speculate that the waning interest is due to economic change, to the passing of the first wave

of reform, or to foreknowledge about the intensity of teacher union reactions. Perhaps an informal network of legislators has purveyed the story that TECAT cost Texas Governor, Mark White, his job. In any case, our research suggests that there is sense to legislators' caution. Although the blame for negative side effects must be widely shared, the realities of teacher testing deny the simplicity of the intended policy—"give a test and eliminate the few teachers with indefensibly weak communication skills." □

¹We disagree with the Commissioner that an additional 8,000 teachers who never showed up to take the test can be claimed as additional successes for the TECAT. Many in this number reflect normal attrition. Based on teacher interview data, some very able teachers left the profession because they were offended by the pettiness of TECAT. This latter contention is consistent with national data showing that teachers leaving the profession are on average higher scoring and better educated than those remaining.

Acknowledgement: Sponsored by a grant from the Office of Educational Research and Improvement, U.S. Department of Education, to the Center for Research on Evaluation, Standards, and Student Testing, OERI-G-86-0003.

-
- American Psychological Association, American Educational Research Association, National Council on Measurement in Education. (1985). *Standards for educational and psychological testing*. Washington, DC: American Psychological Association, Inc.
- Bridges, E.M. (1986). *The incompetent teacher*. Philadelphia: The Falmer Press.
- Kirby, W.N. (1987, April). *A formal team debate: Proposition-Recertification testing of teachers will improve the quality of public schooling*. Proponent in favor of the proposition at the annual meeting of the American Educational Research Association. (Audio recording RA7-24.01). Chicago, IL: Teach'em, Inc., 160 E. Illinois St., 60611.
- Popham, W.J. (1986, April). *Political permutations of standard-setting when the stakes are high*. Paper presented at the annual meeting of the American Educational Research Association, San Francisco.
- Popham, W.J. (1987, April). *A formal team debate: Proposition-Recertification testing of teachers will improve the quality of public schooling*. Proponent in favor of the proposition at the annual meeting of the American Educational Research Association. (Audio recording RA7-24.01). Chicago, IL: Teach'em, Inc., 160 E. Illinois St., 60611.
- Shepard, L.A., Kreitzer, A.E., & Graue, M.E. (1987). *A case study of the Texas teacher test: Technical report*. (OERI-G-86-0003). Los Angeles: Center for Student Testing, Evaluation, and Standards, University of California.
- Sirota and Alper Associates, Inc. (1984, February). *Texas Education Agency: Pilot District Management Process Survey, Statewide Results*. Author.
- Stake, R.E. (1973, October). *Program evaluation, particularly responsive evaluation*. Paper presented at a conference on *Trends in Evaluation*, Goteborg, Sweden.
- The Texas Poll. (1986, February). *Texans continue to rate public schools highly*. Texas A & M University, Public Policy Resources Laboratory. Author.
- Yalow, E.S. (1986, April). *Tempest in a TECAT*. Paper presented at the annual meeting of the American Educational Research Association, San Francisco.