8-19-2000

Education Policy Analysis Archives 08/41

Arizona State University

University of South Florida

Follow this and additional works at: http://scholarcommons.usf.edu/coedu_pub

Part of the Education Commons

Scholar Commons Citation
http://scholarcommons.usf.edu/coedu_pub/299

This Article is brought to you for free and open access by the College of Education at Scholar Commons. It has been accepted for inclusion in College of Education Publications by an authorized administrator of Scholar Commons. For more information, please contact scholarcommons@usf.edu.
The Myth of the Texas Miracle in Education

Walt Haney
Boston College

Abstract:
I summarize the recent history of education reform and statewide testing in Texas, which led to introduction of the Texas Assessment of Academic Skills (TAAS) in 1990-91. A variety of evidence in the late 1990s led a number of observers to conclude that the state of Texas had made near miraculous progress in reducing dropouts and increasing achievement. The passing scores on TAAS tests were arbitrary and discriminatory. Analyses comparing TAAS reading, writing and math scores with one another and with relevant high school grades raise doubts about the reliability and validity of TAAS scores. I discuss problems of missing students and other mirages in Texas enrollment statistics that profoundly affect both reported dropout statistics and test scores. Only 50% of minority students in Texas have been progressing from grade 9 to high school graduation since the initiation of the TAAS testing program. Since about 1982, the rates at which Black and Hispanic
students are required to repeat grade 9 have climbed steadily, such that by the late 1990s, nearly 30% of Black and Hispanic students were "failing" grade 9. Cumulative rates of grade retention in Texas are almost twice as high for Black and Hispanic students as for White students. Some portion of the gains in grade 10 TAAS pass rates are illusory. The numbers of students taking the grade 10 tests who were classified as "in special education" and hence not counted in schools' accountability ratings nearly doubled between 1994 and 1998. A substantial portion of the apparent increases in TAAS pass rates in the 1990s are due to such exclusions. In the opinion of educators in Texas, schools are devoting a huge amount of time and energy preparing students specifically for TAAS, and emphasis on TAAS is hurting more than helping teaching and learning in Texas schools, particularly with at-risk students, and TAAS contributes to retention in grade and dropping out. Five different sources of evidence about rates of high school completion in Texas are compared and contrasted. The review of GED statistics indicated that there was a sharp upturn in numbers of young people taking the GED tests in Texas in the mid-1990s to avoid TAAS. A convergence of evidence indicates that during the 1990s, slightly less than 70% of students in Texas actually graduated from high school. Between 1994 and 1997, TAAS results showed a 20% increase in the percentage of students passing all three exit level TAAS tests (reading, writing and math), but TASP (a college readiness test) results showed a sharp decrease (from 65.2% to 43.3%) in the percentage of students passing all three parts (reading, math, and writing). As measured by performance on the SAT, the academic learning of secondary school students in Texas has not improved since the early 1990s, compared with SAT takers nationally. SAT-Math scores have deteriorated relative to students nationally. The gains on NAEP for Texas fail to confirm the dramatic gains apparent on TAAS. The gains on TAAS and the unbelievable decreases in dropouts during the 1990s are more illusory than real. The Texas "miracle" is more hat than cattle.

Click on items in list below for full text.

- Part 1: Introduction
- Part 2: Recent History of Testing in Texas
- Part 3: Evidence and Boosters of the Myth
- Part 4: Problems with TAAS
- Part 5: Missing Students and Other Mirages
- Part 6: Educators' Views of TAAS
- Part 7: Other Evidence on Education in Texas
- Part 8: Summary and Lessons from the Myth Deflated
- Notes and References
- Appendices

About the Author

Walt Haney
Walt Haney, Ed.D., Professor of Education at Boston College and Senior Research Associate in the Center for the Study of Testing Evaluation and Educational Policy (CSTEEP), specializes in educational evaluation and assessment and educational technology. He has published widely on testing and assessment issues in scholarly journals such as the *Harvard Educational Review*, *Review of Educational Research*, and *Review of Research in Education* and in wide-audience periodicals such as *Educational Leadership*, *Phi Delta Kappan*, *the Chronicle of Higher Education* and the "Washington Post." He has served on the editorial boards of *Educational Measurement: Issues and Practice* and the *American Journal of Education* and on the National Advisory Committee of the ERIC Clearinghouse on Assessment and Evaluation.

ACKNOWLEDGEMENTS

As I worked on this article over a period of more than two years, literally dozens of people helped me in numerous ways. At Boston College, graduate students Cathy Horn, Kelly Shasby, Miguel Ramos and Damtew Teferra helped on specific portions of the work reported here. Damtew Teferra has been especially helpful over a period of nearly two years in tracking down references and other source material and in checking accuracy of data input. Ed Rincon and Terry Hitchcock both helped me more than once as I sought to build the data set on enrollments in Texas public schools over the last quarter century. Among many scholars who have answered repeated questions and kindly provided me with references and reviews of various portions of this article in many different versions, I thank James Hoffman, John Tyler, Jeff Rodamar, Angela Valenzuela, Bob Hauser, Duncan Chaplin, Richard Murnane, Linda McNeil, Dennis Shirley, Anne Wheelock, and Janet Baldwin. Diane Joyce, Lauren McGrath, Anna Geraty, Courtney Danley, and Genia Young helped me, as they do everyone else in our research center, in ways too many to mention here. Also, I thank Jane Hodosi of Market Data Retrieval who, under an extraordinarily short deadline, helped me get the mailing labels that allowed us to carry out two of the surveys recounted here. Thanks too to Chris Patterson for providing me with a great deal of useful information. Three electronic mail correspondents—Audrey Amrein, Craig Bolon and Alein Jehlen also provided me with helpful suggestions and encouragement.

I also wish to express my appreciation to Judge Edward Prado. Though I think he may have erred in his ruling in the GI Forum case (as may be apparent from what is to follow), during the four days I was on the stand in his courtroom, he treated me with attention, respect and good humor. He even had the good sense to tell me simply to "answer the question," when the professor in me launched into discussions of literature on topics on which I was questioned. My wife, Kris, and daughter, Elizabeth, also deserve great appreciation for their tolerance in putting up with work that I told them many times would be done long before now. Thanks also to Gene V Glass who encouraged me to submit this work to *Education Policy Analysis Archives*. As a former editor, I know how hard it sometimes can be to pry manuscripts away from authors who know that there are always other nooks and crannies to explore. Thanks too to nine anonymous reviewers from the *EPAA* Editorial Board who commented generously on a previous version of this article.

More than anyone else, though, I wish to express my appreciation and respect for Al Kauffman. Over the past two years, on several occasions I have cursed him under my breath (and once or twice aloud), for getting me involved with TAAS and education reform in Texas. But after spending more than twice as long on this topic as I ever thought I would, I have developed...
enduring respect for Al, his integrity and good humor, and his quest for truth and justice. I regret that I was not able to complete all of the analyses reported here before the TAAS trial. But it will be a long time before I let Al talk me into working on another case, even if next time he tries to tell me I am not his second choice as an expert witness.

Any errors of fact or interpretation in this report are, of course, despite the enormous help of many good and generous people, entirely my responsibility. No corporations, foundations or anonymous donors have supported the research reported here. But I do owe an enormous debt of gratitude to Boston College for awarding me a sabbatical leave during the 1999-2000 academic year. Without the leave, there is no way I would have been able to complete this research. I did not do what I said I would when I applied for sabbatical leave, but I hope that the work reported here will win me, if not forgiveness, at least tolerance for being distracted from well-intentioned plans.

And on the topic of forgiveness, I am almost certain there are people I should have thanked here but could not remember. Forgive me, please, but I simply had to finish this work before returning to normal academic duties in September.

Copyright 2000 by the Education Policy Analysis Archives

The World Wide Web address for the Education Policy Analysis Archives is epaa.asu.edu

General questions about appropriateness of topics or particular articles may be addressed to the Editor, Gene V Glass, glass@asu.edu or reach him at College of Education, Arizona State University, Tempe, AZ 85287-0211. (602-965-9644). The Commentary Editor is Casey D. Cobb: casey.cobb@unh.edu.

EPAA Editorial Board

Michael W. Apple
University of Wisconsin

Greg Camilli
Rutgers University

John Covaleskie
Northern Michigan University

Alan Davis
University of Colorado, Denver

Sherman Dorn
University of South Florida

Mark E. Fetler
California Commission on Teacher Credentialing

Richard Garlikov
hmwkhelp@scott.net

Thomas F. Green
Syracuse University

Alison I. Griffith
York University

Arlen Gullickson
Western Michigan University

Ernest R. House
University of Colorado

Aimee Howley
Ohio University

Craig B. Howley
Appalachia Educational Laboratory

William Hunter
University of Calgary

Daniel Kallós
Umeå University

Benjamin Levin
University of Manitoba

Thomas Mauhs-Pugh
Green Mountain College

Dewayne Matthews
Western Interstate Commission for Higher Education

William McInerney
Purdue University

Mary McKeown-Moak
MGT of America (Austin, TX)
Les McLean  
University of Toronto

Anne L. Pemberton  
apembert@pen.k12.va.us

Richard C. Richardson  
New York University

Dennis Sayers  
Ann Leavenworth Center for Accelerated Learning

Michael Scriven  
scriven@aol.com

Robert Stonehill  
U.S. Department of Education

Susan Bobbitt Nolen  
University of Washington

Hugh G. Petrie  
SUNY Buffalo

Anthony G. Rud Jr.  
Purdue University

Jay D. Scribner  
University of Texas at Austin

Robert E. Stake  
University of Illinois—UC

David D. Williams  
Brigham Young University

---

EPAA Spanish Language Editorial Board

Associate Editor for Spanish Language  
Roberto Rodríguez Gómez  
Universidad Nacional Autónoma de México

roberto@servidor.unam.mx

---

Adrián Acosta (México)  
Universidad de Guadalajara  
adriano@compuserve.com

J. Félix Angulo Rasco (Spain)  
Universidad de Cádiz  
felix.angulo@uca.es

Teresa Bracho (México)  
Centro de Investigación y Docencia Económica-CIDE  
brachodis1.cide.mx

Alejandro Canales (México)  
Universidad Nacional Autónoma de México  
canales@servidor.unam.mx

Ursula Casanova (U.S.A.)  
Arizona State University  
casanova@asu.edu

José Contreras Domingo  
Universitat de Barcelona  
Jose.Contreras@doe.d5.ub.es

Erwin Epstein (U.S.A.)  
Loyola University of Chicago  
eepstein@luc.edu

Josué González (U.S.A.)  
Arizona State University  
josue@asu.edu

Rollin Kent (México)  
Departamento de Investigación Educativa-DIE/CINVESTAV  
kent@gemtel.com.mx  
kentr@data.net.mx

María Beatriz Luce (Brazil)  
Universidad Federal de Rio Grande do Sul-UFRGS  
lucemb@orion.ufrgs.br

Javier Mendoza Rojas (México)  
Universidad Nacional Autónoma de México  
javiermr@servidor.unam.mx

Marcela Mollis (Argentina)  
Universidad de Buenos Aires  
mollis@filo.uba.ar

Humberto Muñoz García (México)  
Universidad Nacional Autónoma de México  
humberto@servidor.unam.mx

Angel Ignacio Pérez Gómez (Spain)  
Universidad de Málaga  
aiperez@uma.es
Daniel Schugurensky (Argentina-Canada)
OISE/UT, Canada
dschugurensky@oise.utoronto.ca

Jurjo Torres Santomé (Spain)
Universidad de A Coruña
jurjo@udc.es

Simon Schwartzman (Brazil)
Fundação Instituto Brasileiro e Geografia e Estatística
simon@openlink.com.br

Carlos Alberto Torres (U.S.A.)
University of California, Los Angeles
torres@gseis.ucla.edu
1. Introduction

Accountability Narrows Racial Gap in Texas; Expand It
—Editorial headline, USA Today, March 21, 2000, p. 14A

For several years the state of Texas has been widely cited as a model of standards-based education reform. Some have even called recent educational progress in Texas a miracle. Indeed Texas has been cited from west coast to east as a model worthy of emulation by other states. As in the USA Today editorial cited above, the Texas system of educational accountability has even been touted as a model to be followed in federal education legislation. In this article, I review evidence to show that the "miracle" of education reform in Texas is really a myth and illusion. What should be learned from this is not just to be suspicious of the "tall tales" of Texans (as Jeff Rodamar, 2000, put the matter), but that more broadly, we should be cautious in drawing sweeping conclusions about large and complex educational endeavors, based on only one form of evidence, such as test scores. This may seem strange advice coming from one who would call a purported miracle a myth. But as I will explain, even if the Texas approach to education reform is not worthy of emulation elsewhere, there is still something to be learned from Texas about how not to judge the health of education and the progress of education reform elsewhere.

The story of the Texas miracle is reported here in eight parts. Following this introduction, Part 2 provides a summary of recent education history in Texas, with particular focus on how statewide testing has evolved in the Lone Star state over the last two decades into the Texas Assessment of Academic Skills (TAAS) which is now the linchpin of educational accountability in Texas. Part 3 summarizes evidence upon which the Texas tale of success in the 1990s is based, and recounts some of the praise that has been lavished recently on the Texas miracle story. Part 4 summarizes some of the problems with the TAAS tests that make them suspect as sources of evidence about the progress of education in Texas. Part 5 describes the problem of missing students in Texas, and other mirages, reminding us that when trying to interpret summary test results, it is always helpful to pay attention to who is and is not present for the testing. Part 6 summarizes views of educators in Texas about TAAS and teaching and learning in the state. Part 7 reviews other evidence on the status of education in Texas. Finally, the conclusion suggests some broader lessons from this story of the myth of the education "miracle" in Texas—about both the limits of test-based accountability and the need to remember the broad aims of education in a democratic society.

Before reviewing the story of the Texas "miracle," I offer two caveats—one very large, and the other inevitable in any work of limited scope. The big caveat is that approximately two years ago Al Kauffman, Regional Counsel for the Mexican American Legal Defense and Education Fund (MALDEF) persuaded me to serve as an expert witness in a MALDEF lawsuit, GI Forum v. Texas Education Agency, brought against the state of Texas. As a result, I served as one of several expert witnesses for MALDEF in its
effort to prove that the high school graduation test in Texas, the TAAS "exit level" test, has illegal discriminatory impact on Black and Hispanic students. After a trial in the fall of 1999 (which in the press came to be called the "TAAS trial,") the federal judge who heard the case, Edward C. Prado, ruled on January 7, 2000, against MALDEF and for the state of Texas. (Note 1) In essence, Judge Prado ruled that while TAAS does have discriminatory impact on Black and Hispanic students, the use of TAAS to withhold diplomas is not illegal because it is educationally necessary. I am not a legal expert and, hence, in the body of this article will comment only on matters of evidence and facts in the TAAS case. Nonetheless, in appendices to this article, I provide the full text of Judge Prado's ruling, documentation on summary arguments made by the two sides in the case, and my own summary comments on the judge's ruling. (Note 2)

The second caveat is one that is inevitable in any presentation in any medium. One can never tell the whole story. Texas is well known for its size. Hence the territory I try to cover in this article is rather large. To provide some indication of its scope, the TAAS trial lasted for five weeks, and in addition to direct testimony, was based on hundreds of documents submitted by plaintiffs and defendants. Indeed, my personal files on TAAS and the TAAS case occupy six feet of shelf space and several megabytes of computer storage. So, in trying to recount the Texas miracle story and why I think it is a myth, I will have to be somewhat selective. This may seem dangerous since I was on one side of a hard fought legal battle. I make no apologies for that, but want to make it clear simply as fair warning to readers. I leave it to others to judge how fair-minded I have been in recounting this version of the Texas miracle. And one final caution. During preparation for the TAAS trial, Mr. Kauffman, the lead attorney for MALDEF in the TAAS case, several times referred to me as his "Yankee testing expert." While I do now reside in New England, I am actually a native of Texas. So beware the tall tales of Texans.
2. Recent History of Testing in Texas

Texas has seen several waves of education reform over the last several decades. As with reform efforts in many other states, testing has featured prominently in these efforts.

In 1971, in the case of *Rodriguez v. San Antonio Independent School District*, a federal court ruled the system of financing public schools in Texas to be unconstitutional in that it discriminated against students living in poor school districts. Although the U.S. Supreme Court reversed the decision in the *Rodriguez* case in 1973, the case helped spur the Texas legislature into trying to remedy inequities in school finance (Funkhouser, 1990, p. 6). In 1979, the Texas legislature passed the Equal Educational Opportunity Act, which established the first state mandated testing program (Office of Technology Assessment, 1987, p. 271). This was the Texas Assessment of Basic Skills (TABS), a survey-type assessment, without sanctions for test takers, from 1980 to 1985.

Following recommendations of a Select Committee on Education (chaired by H. Ross Perot), in 1984 the Texas legislature passed a comprehensive education reform law mandating the most sweeping changes in education in Texas in 30 years (Funkhouser, 1990, p. 3). Among other things, the law established a statewide curriculum (called the Essential Elements), required students to achieve a score of 70 to pass their high school courses, mandated the "no pass, no play" rule (whereby students could not participate in varsity sports if they did not pass high school courses), required teachers to pass a proficiency test; and mandated changes in the statewide testing program (Funkhouser, 1990). Commenting on the state of education in Texas in the mid-1980s, Harold Hodgkinson observed that "The current Texas school reform is as 'top down' as can be found in the U.S. The costs of operating the system now enacted into law will be severe and the retention rate to high school graduation will likely decrease" (Hodgkinson, 1986).

The 1984 law mandated basic skills testing of students in each odd numbered grade (Funkhouser, 1990, p. 199). The new testing program, called the Texas Educational Assessment of Minimum Skills or TEAMS, was implemented in 1985 and tested students in grades 1, 3, 5, 7, 9 and 11. Under the 1984 law, high school students were required to pass the "exit level" version of TEAMS in order to receive a high school diploma, based on a passing score set by the State Board of Education (Office of Technology Assessment, 1987, pp. 272-75). The TEAMS exit-level tests were given for the first time in October 1985 to approximately 190,000 eleventh graders. Eighty-eight percent of students passed the math portion of TEAMS; 91 percent passed the English language arts portion; and 85 percent passed both. Students who failed either portion of TEAMS had an opportunity to retake the tests in May 1986. The majority of students, who had failed in the fall, passed the spring retest (Funkhouser, 1990, pp. 199-201).

In Fall 1990, changes in state law required the implementation of a new "criterion-referenced" testing program, the Texas Assessment of Academic Skills (TAAS) and also established end-of-course tests for selected high school course subjects. As compared with TEAMS, TAAS was intended to shift the focus of assessment from "minimum skills to academic skills" and to test "higher-order thinking skills and problem
solving ability." (TEA, 1997, p. 1). The TAAS is developed for Texas by National Computer Systems, which subcontracts for portions of work to Harcourt Brace Educational Measurement (for item development) and Measurement Incorporated (for scoring of the open-ended portions of the TAAS). TAAS was administered to students in grades 3, 5, 7, and 11 in Fall of 1990 and 1991.

Results of the fall 1990 tryout of TAAS showed that the new tests were much more difficult than the TEAMS tests had been. Table 2.1 shows results from the Fall 1990 grade 11 field test of TAAS. These results made clear that if the passing score on TEAMS (70% correct) was maintained for TAAS, passing rates would fall from the 80-90% range seen on TEAMS to the 40-60% range on TAAS (with pass rates for Black and Hispanic students on the math portion of TAAS falling to the 27-33% range).

Table 2.1
Possible Passing Scores Based on Texas Assessment of Academic Skills (TAAS) Field Test Results Exit (11) (1990)

<table>
<thead>
<tr>
<th>Mathematics (Total possible score is 60 items correct)</th>
<th>Projected Percent Passing</th>
</tr>
</thead>
<tbody>
<tr>
<td>Number of items</td>
<td>Percent of Items</td>
</tr>
<tr>
<td>36</td>
<td>60%</td>
</tr>
<tr>
<td>42</td>
<td>70%</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Reading (Total possible score is 48 items correct)</th>
<th>Projected Percent Passing</th>
</tr>
</thead>
<tbody>
<tr>
<td>Number of items</td>
<td>Percent of Items</td>
</tr>
<tr>
<td>29</td>
<td>60%</td>
</tr>
<tr>
<td>34</td>
<td>71%</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Writing (Total possible score is 40 items correct)</th>
<th>Projected Percent Passing</th>
</tr>
</thead>
<tbody>
<tr>
<td>Number of items</td>
<td>Percent of Items</td>
</tr>
<tr>
<td>24</td>
<td>60%</td>
</tr>
<tr>
<td>28</td>
<td>70%</td>
</tr>
</tbody>
</table>


The 1992-93 school year was a time of transition for statewide testing in Texas with some grades being tested in the fall and some in the spring. In the Spring of 1994, the TAAS reading and mathematics assessments were administered to students in grades 3, 4, 5, 6, 7, 8, and 10; and the TAAS writing tests were administered at Grades 4, 8, and 10. If students do not pass the grade 10 or exit level TAAS, they may continue taking
portions they have not yet passed during grades 11 and 12. Since 1994, the TAAS Reading, Mathematics and Writing tests have consistently been administered to students in grades 4, 8 and 10 in the spring of each year.

In addition to being used to help ensure student learning, TAAS results are also used to hold schools and school systems "accountable" for student learning. By state law, the State Board of Education is mandated to rate the performance of schools and school districts according to a set of "academic excellence indicators," including TAAS results, dropout rates and student attendance rates (TEA, 1997, p. 159). State law also prescribes that student performance data be disaggregated by ethnicity and socioeconomic status. The performance rating system holds that school performance is not acceptable if the performance of all subgroups is not acceptable. Based primarily on percentage of students passing each of the TAAS tests, the more than 6,000 schools in Texas have been rated since 1994 as "exemplary," "recognized," "acceptable" or "unacceptable."

TAAS passing standards [for schools' performance ratings] . . . are based on the passing rates for all students and the disaggregated rates for four student groups: African American, Hispanic, White, and Economically Disadvantaged. Of the four categories, only the exemplary rating has had a consistent passing standard, requiring at least 90 percent of all students and each student group to pass each subject area. The recognized rating has increased from at least 65 percent of students passing in 1994 to a current 70 percent, the acceptable rating has gone from at least 25 percent passing to 30 percent, and the low-performing rating has gone from at least 25 percent passing to 30 percent, and the low-performing rating from less than 25 percent to less than 30 percent. (Gordon & Reese, 1997, p. 347-480)

Schools are eligible for cash awards for high ratings; and if they are rated as low performing twice in a row, they are subject to sanctions from the Texas Education Agency, including possible closure.

In short, over the past decade TAAS has become an extremely high stakes test for students, educators and schools in the state of Texas. If students do not pass all three portions of the exit level version of TAAS (reading, math and writing), they cannot graduate from high school, regardless of grades in their high school courses. And schools' reputations, funding and their continued existence depend on students' performance on TAAS. (Note 3)

Before summarizing TAAS results in the 1990s, it is useful to describe the tests themselves. The focus of test-based accountability in Texas is on the TAAS tests of reading, mathematics and writing (there are also TAAS tests of social studies and science and end-of course tests in some high school subjects). The TAAS tests are mostly multiple-choice in format. The numbers of questions on the TAAS tests varies somewhat across grade level versions, but the grade 10 (or exit level) versions contain 48 reading questions, 60 math questions and 40 writing questions. The TAAS writing test also includes an open-ended question to which students must write their answers. The written composition portion of the TAAS writing test is scored on a 4-point scale (released versions of the TAAS tests are available at www.tea.state.tx.us/student.assessment/release.htm).

Finally, I should mention that though several observers have described the TAAS tests as criterion-referenced, traditional norm-referenced test construction techniques (such as screening of candidate items in terms of item discrimination) have been used in their construction. Also it is clear that the TAAS tests have so few items that they cannot be used to yield reliable scores below the aggregate reading, math and writing
levels—and as we will see, there is ample cause to doubt their reliability and validity even at these aggregate levels. Moreover, as will be explained, the passing scores on the TAAS test were set without any reference to performance criteria external to TAAS, but only after review of group performance on TAAS—in effect a norm-referenced rather than criterion-referenced comparison.

As mentioned, by law the Texas State Board of Education was required to set passing scores on the TAAS tests (or as legislative language put it, "determine the level of performance considered to be satisfactory," TEA, 1997, p. 157). Here is how the Texas Student Assessment Program Technical Digest describes the evolution of the TAAS "passing standard":

In 1990 the State Board of Education set minimum expectations as equivalent to 70% of the multiple-choice items correct on the fall 1990 test and a score of at least 2 on the written composition. The 70%-equivalent standard was in effect beginning with the 1991-1992 school year. The 1990-1991 school year served as a transition from the previous assessment program, The Texas Assessment of Minimum Skills (TEAMS). The SBOE set the interim minimum expectations standard at 65% of the multiple-choice items correct for Grades 3, 3-Spanish, and 5, and 60% of the items correct for grades 7, 9 and exit level. A student also had to score at least 2 on the written composition to meet minimum expectations on the writing test. (TEA, 1997, p. 28)

So, since 1992 the passing scores on the TAAS exit level tests (reading, writing and math) have been set at a level equivalent to the 70% of items correct on Fall 1990 form of the tests. As new forms of the tests were used in subsequent years, analysts used test-equating methods to try to make passing scores on the new forms equivalent to 70% correct on the 1990 forms. (Note 4)
3. Evidence and Boosters of the Myth

Given the consequences attached to performance on TAAS, it is not surprising that this test has had major impact on education in Texas. At first glance, this impact appears to have been largely positive; and it is evidence of the apparent positive impact of TAAS, and the Texas system of school accountability, that has helped give rise to the "miracle" story of education reform in Texas over the last decade.

Four kinds of evidence seem to have been most widely cited as indicative of major improvements in education in Texas, namely: 1) sharp increases in the overall pass rates on TAAS during the 1990s; 2) apparent decreases in the achievement gap between White and minority students in Texas (again based on TAAS scores); 3) seemingly decreasing rates of students dropping out of school before high school graduation; and, 4) apparent confirmation of TAAS gains by results on the National Assessment of Educational Progress (NAEP).

3.1 Improved results on TAAS

The main evidence contributing to the perception of dramatic educational gains in Texas during the 1990s (what the March 21, 2000 USA Today editorial called "widespread improvement in student achievement") seems to have been sharp increases in passing rates on the TAAS. TAAS was introduced in Texas in 1990-91, and, as recounted previously, was administered at somewhat varied grades (and seasons) during the early 1990s. In several publications, the TEA has presented TAAS pass rates aggregated across different grades. Inasmuch as this sort of aggregation may obscure as much as it reveals, here I present results mainly for grade 10 TAAS testing.

Table 3.1 (and corresponding Figure 3.1) shows the results on the grade 10 TAAS test from 1994 to 1998.

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>TAAS Reading</td>
<td>76%</td>
<td>76%</td>
<td>81%</td>
<td>86%</td>
<td>88%</td>
</tr>
<tr>
<td>TAAS Math</td>
<td>57%</td>
<td>59%</td>
<td>65%</td>
<td>72%</td>
<td>78%</td>
</tr>
<tr>
<td>TAAS Writing</td>
<td>81%</td>
<td>86%</td>
<td>85%</td>
<td>88%</td>
<td>89%</td>
</tr>
<tr>
<td>TAAS All Tests</td>
<td>52%</td>
<td>54%</td>
<td>60%</td>
<td>67%</td>
<td>72%</td>
</tr>
</tbody>
</table>
As can be seen from these data, grade 10 TAAS results show a pattern of steady improvement from 1994 through 1998, with the percentage of students passing the TAAS reading test rising from 76% to 88%; the percentage passing the TAAS math test rising from 57% to 78%; and the corresponding increase for the TAAS writing test going from 81% to 89%. The percentage of grade 10 students passing all three tests increased from 52% in 1994 to 72% in 1998.

3.2 Decrease in Race Gap in Test Scores

Even as test scores were improving overall, the gaps in achievement between White and nonwhite students (specifically Black and Hispanic students) appeared to have been narrowing. The *USA Today* editorial (3/21/2000) reported that "Texas is one of the few states that has narrowed its racial learning gap." Figure 3.2 and Table 3.2 show how the "racial learning gap" appears to have narrowed on the grade 10 TAAS tests (for economy of presentation here, I do not show results separately for the reading, writing, and math tests, but only the percentages of grade 10 students passing all three tests).

### Table 3.2

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Black</td>
<td>29%</td>
<td>32%</td>
<td>38%</td>
<td>48%</td>
<td>55%</td>
</tr>
</tbody>
</table>
Hispanic 35% 37% 44% 52% 59%
White 67% 70% 74% 81% 85%

Source: Selected State AEIS Data: A Multi-Year History: www.tea.state.tx.us/student.assessment/swresult/gd10sp98.htm

As can be seen, in 1994 there was a huge disparity in the grade 10 pass rates for Black and Hispanic students as compared with White students. The 1994 White pass rate of 67% was 38 points higher than the Black pass rate of 29%; and 32 points more than the Hispanic rate of 35%. In other words, in 1994, White students were passing the grade 10 TAAS tests at about double the rate of Black and Hispanic students. This gap was just about what might have been predicted based on the 1990 field test results (see Table 2.1). By 1998, the White grade 10 pass rate had climbed 18 points to 85%. But the Black and Hispanic pass rates had climbed even more, 26 and 24 points respectively. So in a period of just five years, the race gaps had been reduced from 38 to 30 percentage points for Whites and Blacks and from 32 to 26 for Whites compared with Hispanic tenth grade students. Or in other words, minorities had increased their rate of passing grade 10 TAAS tests from less than 50% of the White pass rate to two-thirds of the White pass rate in just four years.

3.3 Decreases in Dropout Rates

If the dramatic gains in grade 10 pass rates overall and substantial decreases in the "racial learning gap" were not sufficiently remarkable, official TEA statistics indicated that over the same interval high school dropout rates were also declining.

Table 3.3
Texas Annual Dropout Rate, Grades 7-12 1994-1998
### Table 3.3

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>All Students</td>
<td>2.8%</td>
<td>2.6%</td>
<td>1.8%</td>
<td>1.8%</td>
<td>1.6%</td>
</tr>
<tr>
<td>Black</td>
<td>3.6%</td>
<td>3.2%</td>
<td>2.3%</td>
<td>2.3%</td>
<td>2.0%</td>
</tr>
<tr>
<td>Hispanic</td>
<td>4.2%</td>
<td>3.9%</td>
<td>2.7%</td>
<td>2.5%</td>
<td>2.3%</td>
</tr>
<tr>
<td>White</td>
<td>1.7%</td>
<td>1.5%</td>
<td>1.2%</td>
<td>1.1%</td>
<td>1.0%</td>
</tr>
</tbody>
</table>

Source: Selected State AEIS Data: Five Year History
www.tea.state.tx.us/perfreport/aeis/hist/state.html

As shown in Table 3.3, TEA data indicated that between 1994 and 1998, even as pass rates on the TAAS were increasing among grade 10 students, dropout rates were decreasing not just among secondary students overall, but also for each of the three race groups for which data were disaggregated. In short, what appeared to be happening in Texas schools in the 1990s truly did seem to be a miracle.

As Peter Schrag has recently written: "Some of Texas's claims are so striking they border on the incredible. The state's official numbers show that even as TAAS scores were going up, dropout rates were cut from an annual 6.1 percent in 1989-90 to 1.6 percent last year. If ever there was a case of something being too good to be true, this is it" (Schrag, 2000). But before reviewing the doubts of Schrag and others, let me recap one additional source of evidence that seemed to confirm the miracle story.

#### 3.4 NAEP Results for Texas

Anyone even remotely familiar with recent education history of the United States must view with some skepticism the meaningfulness of the almost inevitable increases in performance that follow introduction of a new testing program. When a new testing program is introduced, students and teachers have little familiarity with the specifics of the new tests. But after a few years, they become familiar with the style and format of the tests and students can be coached specifically for the test in question. Hence, performance—or at least average test scores—almost inevitably increases.

That students can be successfully coached for particular tests has been well known among education researchers for decades. As far back as 1927, Gilmore, for example, reported that students could be coached on Otis group intelligence tests "to the point of increasing their standing and score in intelligence tests even in the case of the material used in coaching being only similar and not identical with that of the basic test" (Gilmore, 1927, p. 321). Indeed what happens when students are coached for a specific test has come to called the "saw tooth" phenomenon because of the regular pattern in which scores steadily rise following introduction of a new testing program, only to fall dramatically when a different test is introduced (Linn, 2000, p. 7).

The phenomenon of falsely inflated test scores was brought to wide public attention in the late 1980s and early 1990s because of publicity for what came to be known as the "Lake Wobegon" phenomenon in test results. Lake Wobegon is the mythical town in Minnesota popularized by Garrison Keillor in his National Public Radio program "A Prairie Home Companion." It is the town where "all the women are strong, all the men are good looking, and all the children are above average." In the late 1980s it was discovered that Lake Wobegon seemed to have invaded the nation's schools. For according to a 1987 report by John Cannell, the vast majority of school districts and all states were scoring above average on nationally normed standardized tests (Cannell, 1987). Since it is logically impossible for all of any population to be above average on a
single measure, it was clear that something was amiss, that something about nationally normed standardized tests or their use had been leading to false inferences about the status of learning in the nation's schools.

Cannell was a physician by training and not a specialist in education or education research. His original (1987) report was published by "Friends for Education," the foundation he established to promote accountability in education. A revised version of Cannell's report was published in the Summer 1988 issue of *Educational Measurement: Issues and Practice* (Cannell, 1988) together with responses and commentary from representatives of major test publishers and officials of the U.S Department of Education (Phillips and Finn, 1988; Drahozal and Frisbie, 1988; Lenke and Keene, 1988; Williams, 1988; Qualls-Payne, 1988; Stonehill, 1988). Cannell's charges regarding misleading test results were hotly debated in this and other forums. Some people doubted whether the Lake Wobegon phenomenon was real (that is, whether large majorities of states, schools and districts were in fact scoring above average on the national norms of the tests), while most observers accepted the reality of the phenomenon but disputed what caused it. Among the causes suggested and debated were problems in the original norming of the tests, outdated norms, lack of test security, manipulation of populations of students tested, artificial statistical manipulation of test results, and teachers and schools teaching to the tests, either purposely or inadvertently.

The publicity surrounding the Lake Wobegon phenomenon was sufficiently widespread that the U.S. Department of Education funded researchers at the Center for Research on Evaluation, Standards and Student Testing (CRESST) to investigate. On the basis of a survey of state directors of testing, Shepard (1989) concluded that the conditions for inflated test results—such as high stakes being pinned on test results, efforts to align curricula to the tests, and direct teaching to the tests—existed in virtually all of the states. And on the basis of an analysis of up to three years of test results from 35 states from which they were available, Linn, Graue and Sanders (1989) essentially confirmed Cannell's basic finding that test results across the nation were implausibly inflated—Lake Wobegon had invaded the nation's schools. For instance, they found that "for grades 1 through 6, the percentage of students scoring above the national median in mathematics ranges from a low of 58% in grade 4 for the 1985 school year to a high of 71% in grade 2 for the 1987-88 school year . . . " (p. 8). Linn, Graue and Sanders concluded that the use of old norms was one cause of the abundance of "above average scores" (p. 23), but also pointed out that in situations in which the same form of a test is used year after year, "increased familiarity with a particular form of a test" (p.24) likely contributed to inflated scores.

The practice of using a single form of a test year after year poses a logical threat to making inferences about the larger domain of achievement. Scores may be raised by focusing narrowly on the test objectives without improving achievement across the broader domain that the test objectives are intended to represent. Worse still, practice on nearly identical or even the actual items that appear on a test may be given. But as Dyer aptly noted some years ago, "if you use the test exercises as an instrument of teaching you destroy the usefulness of the test as an instrument for measuring the effects of teaching (Dyer, 1973, p. 89)." (Linn, Graue and Sanders, 1989, p. 25).

The problem was illustrated even more clearly in a subsequent study reported by Koretz, Linn, Dunbar & Shepard (1991), which compared test results on one "high-stakes" test, used for several years in a large urban school district, with those on a
comparable test that had not been used in that district for several years. They found that performance on the regularly used high-stakes test did not generalize to other tests for which students had not been specifically coached, and again commented that "students in this district are prepared for high-stakes testing in ways that boost scores . . . substantially more than actual achievement in domains that the tests are intended to measure" (p. 2). To put the matter bluntly, teaching to a particular test undermines the validity of test results as measures of more general learning.

While education researchers were essentially confirming Cannell's initial charges, the intrepid physician was continuing his own investigations. In late summer 1989, Cannell released a new report entitled The "Lake Wobegon" Report: How Public Educators Cheat on Standardized Achievement Tests. This time Cannell presented new instances of the Lake Wobegon phenomenon and a variety of evidence of outright fraud in school testing programs, including a sampling of testimony from teachers concerned about cheating on tests. After presenting results of his own survey of test security in the 50 states (concluding that security is generally so lax as to invite cheating), Cannell outlined methods to help people detect whether cheating is going on in their school districts, and "inexpensive steps" to help prevent it.

More recently Koretz and Barron (1998; RAND, 1999) of the RAND Corporation investigated the validity of dramatic gains on Kentucky's high stakes statewide tests. Like Texas, Kentucky had adopted policies to hold schools and teachers accountable for student performance on statewide tests. During the first four years of the program, Kentucky students showed dramatic improvements on the state tests. What Koretz and Barron sought to assess was the validity of the Kentucky test gains by comparing them with Kentucky student performance on comparable tests, specifically the National Assessment of Educational Progress (NAEP) and the American College Testing Program (ACT) college admissions tests. What they found was that the dramatic gains on the Kentucky test between 1992 and 1996 were simply not reflected in NAEP and ACT scores. They concluded that the Kentucky test scores "have been inflated and are therefore not a meaningful indicator of increased learning" (RAND, 1999).

Even before the release of the report showing inflated test scores in Kentucky, anyone familiar with the Lake Wobegon phenomenon, widely publicized in the late 1980s and early 1990s, had to view the dramatic gains reported on TAAS in Texas in the 1990s with considerable skepticism. Were the gains on TAAS indicative of real gains in student learning, or just another instance of artificially inflated test scores?

In 1997, results from the 1996 the National Assessment of Educational Progress (NAEP) in mathematics were released. The 1996 NAEP results showed that among the states participating in the state-level portion of the math assessment, Texas showed the greatest gains in percentages of fourth graders scoring at the proficient or advanced levels. Between 1992 and 1996, the percentage of Texas fourth grades scoring at these levels had increased from 15% to 25%. The same NAEP results also showed North Carolina to have posted unusually large gains at the grade 8 level, with the percentages of eighth graders in North Carolina scoring at the proficient or advanced levels improving from 9% in 1990 to 20% in 1996. (Reese et al., 1997)

Putting aside for the moment that the 1996 NAEP results also showed that math achievement in these two states was no better (and in some cases worse) than the national average, these findings led to considerable publicity for the apparent success of education reform in these two states. The apparent gains in math, for example, led the National Education Goals Panel in 1997 to identify Texas and North Carolina as having made unusual progress in achieving the National Education Goals.
3.5 Plaudits for the Texas Miracle

In Spring 1998, Tyce Palmaffy published an article titled "The Gold Star State: How Texas jumped to the head of the class in elementary school achievement." Citing both 1996 NAEP results and TAAS score increases, Palmaffy praised Texas for being in the vanguard of "an accountability movement sweeping the states" (not surprisingly he also mentioned North Carolina and Kentucky). Regarding TAAS, Palmaffy reported "In 1994, barely half of Texas students passed the TAAS math exam. By last year, the proportion had climbed to 80 percent. What's more, the share of black and Hispanic children who passed the test doubled during that time to 64 percent and 72 percent respectively." Palmaffy's article, published in a Heritage Foundation journal, also included testimonials for the Texas success story from divergent vantage points. Kati Haycock, "director of the Education Trust, a Washington D.C.-based organization devoted to improving educational opportunities for low-income children" was quoted as touting Texas as "a real model for other states to follow." The article also referred to "researcher Heidi Glidden of the American Federation of Teachers union" as praising the sort of education accountability system used in Texas.

Meanwhile, the National Education Goals Panel had "commissioned Dr. David Grissmer, an education researcher with the RAND Corporation, to conduct an analysis of education reforms in both states [Texas and North Carolina] to determine that the improvements were indeed significant and to seek to identify the factors that could and could not account for their progress" (Grissmer & Flanagan, 1998, p. i). The National Education Goals Panel released the Grissmer/Flanagan report in November 1998. Without trying to recap or critique the Grissmer/Flanagan report here, let me simply summarize how it was conveyed to the outside world. The report was released November 5, 1998 with a press release titled "North Carolina and Texas Recognized as Models for Boosting Student Achievement." The first paragraph of the press release read:

(WASHINGTON, D.C.) A new study that both belies conventional wisdom about problems in K-12 education and illuminates some approaches for solving them points to the extraordinarily successful policies of two states North Carolina and Texas as models for reform throughout the nation. (NEGP, 11/5/98)

After quotes from North Carolina Governor Jim Hunt and Texas Governor George W. Bush, the press release went on to summarize the Grissmer/Flanagan findings. The researchers found that "several factors commonly associated with student achievement, such as real per pupil spending, teacher pupil ratios, teachers with advanced degrees, and experience level of teachers—are not adequate for explaining the test score gains." (National Education Goals Panel, November 5, 1998, p. 1). The press release explained that, instead, Grissmer and Flanagan attributed the achievement gains in Texas and North Carolina to three broad factors common to the two states (business leadership, political leadership, consistent reform agendas) and seven educational policies (adopting statewide standards by grade for clear teaching, holding all students to the same standards, linking statewide assessments to academic standards, creating accountability systems with benefits and consequences for results, increasing local control and flexibility for administrators and teachers, providing test scores and feedback via computer for continuous improvement, and shifting resources to schools with more disadvantaged students).
Grissmer and Flanagan (1998) did not explain how they had determined that these were the factors behind the apparent achievement gains in Texas and North Carolina; but whatever the case, this 1998 report from the National Education Goals Panel, coupled with the sort of diverse support for the Texas model education accountability system cited by Palmaffy, seemed to certify the apparent miracle of education reform in Texas. The success of education reform in Texas was being heralded by observers as diverse as Palmaffy (of the Heritage Foundation), Haycock (head of an organization dedicated to improving the educational opportunities of low-income children), and Glidden (a researcher with one of the nation's largest teachers unions). The Grissmer/Flanagan report seemed to be the clincher. Here was a report from a bipartisan national group (the National Education Goals Panel), prepared by a Ph.D. researcher from a prestigious research organization, the RAND Corporation, that straight out said, "The analysis confirms that gains in academic achievement in both states are significant and sustained. North Carolina and Texas posted the largest average gains in student scores on tests of the National Assessment of Educational Progress (NAEP) administered between 1990 and 1997. These results are mirrored in state assessments during the same period, and there is evidence of the scores of disadvantaged students improving more rapidly than those of advantaged students" (Grissmer & Flanagan, 1998, p. i). Few people seemed to notice that the Grissmer & Flanagan report was not actually published by RAND.

Nonetheless, the report from the National Education Goals Panel seemed to certify the seeming miracle of education reform in Texas. Subsequently, the story of the Texas miracle has been circulated far and wide. Without trying to document all of the stories on the Texas miracle I have seen, let me mention here just two examples. On June 10, 1999, the Boston Globe ran a front-page story headlined "Embarrassed into success: Texas school experience may hold lessons for Massachusetts" (Daley, 1999). And on March 21, 2000, in the editorial cited at the start of this article, USA Today, in urging the U.S. Senate to adopt a Texas-style school accountability system for the $8 billion Title I program providing federal aid to poor schools, the editors cited "Texas-size school success" in the Lone Star state. In an apparent reference to 1996 NAEP results, the editorial cited the Education Trust as the source of evidence about gains in Texas on 1996 math tests administered nationally.
4. Problems with TAAS

Two years ago when I agreed to help MALDEF on the TAAS case, I had no way of foreseeing the extent to which education reform in Texas would come to be touted as a model to be emulated elsewhere. Nonetheless, as I studied what had been happening with TAAS in Texas, I quickly came to think otherwise. Before summarizing what I think is wrong with TAAS and how it is being misused in Texas, I should mention that some of what I recount in the remainder of this article is based on two unpublished reports that I prepared in connection with the TAAS case—a preliminary report in December 1998, and supplementary report in July 1999 (Haney, 1998; 1999). However, it also draws on additional evidence acquired and analyses undertaken since completion of the supplementary report in summer 1999.

The problems with TAAS and the way it is being used in Texas may be summarized under five sub-headings: 1) the TAAS is having a continuing adverse impact on Black and Hispanic students; 2) the use of the TAAS test in isolation to control award of high school diplomas is contrary to professional standards concerning test use; 3) the passing score on TAAS is arbitrary and discriminatory; 4) a variety of evidence casts doubt on the validity of TAAS scores; and 5) more appropriate use of test results would have more validity and less adverse impact.

4.1 Adverse impact

In previous research and law, three standards have been recognized for determining whether observed differences constitute discriminatory disparate impact: 1) the 80 percent (or four-fifths) rule; 2) tests of the statistical significance of observed differences; 3) and evaluation of the practical significance of differences. The "80 percent" or four-fifths rule refers to a provision of the 1978 Uniform Guidelines on Employee Selection Procedures (43 F.R. No. 166, 38290-38296, 1978) which reads:

Sec. 6D. Adverse impact and the "four-fifths rule." A selection rate for any race, sex or ethnic group which is less than four-fifths (or eighty percent) of the rate for the group with the highest rate will be generally regarded by Federal enforcement agencies as evidence of adverse impact, while a greater than four-fifths rate will generally not be regarded by Federal enforcement agencies as evidence of adverse impact. (As quoted in Fienberg, 1989, p. 91).

As a result of its standing in federal regulations, the 80 percent rule as a test of adverse or disparate impact has been widely recognized. Nonetheless, simple differences in percentage rates have some undesirable properties. The simple difference, for example "is inevitably small when the two percentages are close to zero" (David H. Kaye and David A. Freedman, Reference guide on statistics, Federal Judicial Center, 1994). Hence, most observers and considerable case law now hold that in assessing disparate impact, it
is important to apply not just the 80% or four-fifths rule but also to consider the practical and statistical significance of differences in selection or pass rates (Fienberg, 1989; Kaye & Freedman, 1994; see also, Office of Civil Rights, 1999). In previous reports regarding the TAAS case (Haney, 1998; 1999), I applied these three tests of adverse impact to a variety of TAAS results. However, for economy of presentation here, I provide only illustrative results.

**Eighty Percent or Four-Fifths Rule.** To apply this test of adverse impact, we simply multiply the pass rates on TAAS for White students by 80% and check to see whether the pass rates for Blacks and Hispanics fall below these levels. Table 4.1 presents the application of the 80% rule to the TAAS results previously presented in Table 3.2 above. As can be seen, even though grade 10 pass rates for all three TAAS tests for Black and Hispanics have improved between 1994 and 1998, these pass rates still lag below 80% of the White pass rates. According to this standard of adverse impact, the TAAS grade 10 tests continue to show adverse impact on Black and Hispanic students. (Note 5)

**Table 4.1**

**Eighty Percent Rule and TAAS Grade 10 Pass Rates: Percent Passing All Tests by Race 1994-1998 All Students Not in Special Education**

(Does Not Include Year-Round Education Results)

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>White</td>
<td>67%</td>
<td>70%</td>
<td>74%</td>
<td>81%</td>
<td>85%</td>
</tr>
<tr>
<td>White*80%</td>
<td>53.6%</td>
<td>56.0%</td>
<td>59.2%</td>
<td>64.8%</td>
<td>68.0%</td>
</tr>
<tr>
<td>Black</td>
<td>29%</td>
<td>32%</td>
<td>38%</td>
<td>48%</td>
<td>55%</td>
</tr>
<tr>
<td>Hispanic</td>
<td>35%</td>
<td>37%</td>
<td>44%</td>
<td>52%</td>
<td>59%</td>
</tr>
</tbody>
</table>

Source: Selected State AEIS Data: A Multi-Year History

**Statistical Significance of Differences in Pass Rates.** As mentioned, comparisons of simple percentages passing have some weaknesses from a statistical point of view. For example, differences in pass rates, particularly if small numbers of examinees are involved, may result from random variation in the particular sample of candidates who take an examination in a particular year. To check against this possibility, a second kind of standard for evaluating discriminatory disparate impact is generally employed; namely, a test of the statistical significance of observed differences. A test of statistical significance is used to assess the probability that a particular outcome (such as differences in proportions passing a test) might have occurred simply by chance or random sampling.

The obvious statistical significance test to apply in a case such as that of proportions of candidates passing the TAAS is the test of the difference in proportions of two populations. As explained in most statistics textbooks, such as Paul Hoel's *Introduction to mathematical statistics* (1971, pp. 134-137), if \( p_1 \) and \( p_2 \) refer to the proportions of successes in two samples, \( q_1 \) and \( q_2 \) refer to the proportions of failures in the two samples, and \( n_1 \) and \( n_2 \) refer to the sizes of the samples, the standard error of the difference in proportions is calculated as follows:
Using this formula we may calculate the standard error of the difference in proportions for each comparison we wish to make and then divide the standard error of the difference into the observed difference to calculate the number of standard errors equivalent to the observed difference. Table 4.2 shows the results of such calculations for the Spring 1998 TAAS results.

### Table 4.2
#### Statistical Significance of Differences in 1998 Grade 10 Pass Rates

<table>
<thead>
<tr>
<th></th>
<th>TAAS Reading</th>
<th>TAAS Math</th>
<th>TAAS Writing</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>No. Tested</td>
<td>% Pass</td>
<td>No. Tested</td>
</tr>
<tr>
<td><strong>Black</strong></td>
<td>26790</td>
<td>81%</td>
<td>27434</td>
</tr>
<tr>
<td><strong>Hispanic</strong></td>
<td>70666</td>
<td>79%</td>
<td>71747</td>
</tr>
<tr>
<td><strong>White</strong></td>
<td>108887</td>
<td>95%</td>
<td>109595</td>
</tr>
</tbody>
</table>

Source: TAAS Summary Report—Test Performance All Students Not In Special Ed. Grade 10—Exit Level Report Date April 98 Date of Testing: March 1998 (www.tea.state.tx.us/student.assessment/results/sum98/g xen98.htm)

<table>
<thead>
<tr>
<th></th>
<th>White-Black Differences</th>
<th>White-Hispanic Differences</th>
</tr>
</thead>
<tbody>
<tr>
<td>SE of difference</td>
<td>0.0025</td>
<td>0.0017</td>
</tr>
<tr>
<td>Obs’d Difference</td>
<td>14%</td>
<td>16%</td>
</tr>
<tr>
<td>Obs’d Diff/SE</td>
<td>56.312</td>
<td>95.894</td>
</tr>
<tr>
<td></td>
<td>0.0031</td>
<td>0.002</td>
</tr>
<tr>
<td>Obs’d Difference</td>
<td>27%</td>
<td>21%</td>
</tr>
<tr>
<td>Obs’d Diff/SE</td>
<td>86.982</td>
<td>104.41</td>
</tr>
<tr>
<td></td>
<td>0.0023</td>
<td>0.0016</td>
</tr>
<tr>
<td>Obs’d Difference</td>
<td>12%</td>
<td>14%</td>
</tr>
<tr>
<td>Obs’d Diff/SE</td>
<td>51.721</td>
<td>89.503</td>
</tr>
</tbody>
</table>

As can be seen from Table 4.2, the differences in pass rates for both White-Black and White-Hispanic comparisons are easily statistically significant, with observed differences equivalent to some fifty to over 100 standard errors. (Other statistical tests on TAAS results also yield results of this magnitude; see Haney, 1998; 1999).

Practical significance of observed differences

What of the practical significance of the observed differences in the 1998 grade 10 TAAS pass rates? Later in this report, I discuss the apparent consequences of the TAAS for grade retention and dropping out of school, but for the moment let us simply examine the numbers of students involved in the differential pass rates.

On the TAAS writing test in 1998, 96% of White students passed, 84% of Black students and 82% of Hispanic students. While these differences do not exceed the 80% rule (96%*0.80 = 76.8%), let us consider the numbers of students involved. Specifically
we may consider the numbers of Black and Hispanic students who would have passed the 1998 grade 10 writing test had the passing rates for Black and Hispanic students been the same as that for White students. These numbers are approximately 3,200 Black students and 9,900 Hispanic students, for a total of about 13,000 (comparable calculations show that on the TAAS math for 1998, about 22,000 more Black and Hispanic students would have passed had their pass rates been the same as for White students). Do the differential results on the 1998 grade 10 TAAS writing test, on which approximately 13,000 more Black and Hispanic students failed than would have been the case had the Black and Hispanic pass rates been the same as that of White students, constitute practical adverse impact? Do the differential results on all of the 1998 grade 10 TAAS tests, on which close to 34,000 more Black and Hispanic students failed (10,700 Black and 23,200 Hispanic students) than would have been the case had the Black and Hispanic pass rates been the same as that for White students constitute practical adverse impact? The answer, especially when results are also suspect under both the 80% rule and tests of statistical significance, seems clear, at least to me. A test that leads to failure for tens of thousands more minority than non-minority students, had they had equivalent passing rates, surely has practical adverse impact. Hence, the validity and educational necessity of such a test deserve close scrutiny.

Before turning to those issues, however, I should mention that in his opinion in the TAAS case on January 7, 2000, Judge Prado ruled that "Plaintiffs have made a prima facie showing of significant adverse impact" (p. 23, though it should be added that the opinion has a discussion of disparate impact in two places, pp.15-17 and 20-23)

4.2 TAAS Use in Isolation Violates Professional Standards

The use of TAAS scores in isolation to control award of high school diplomas (or for that matter use of any test results alone to make high stakes decisions about individuals or institutions) is contrary both to professional standards regarding testing and to sound professional practice.

The standards to which I refer are the Standards for Educational and Psychological Testing published by the American Educational Research Association (AERA), the American Psychological Association (APA) and the National Council on Measurement in Education (NCME). These standards have been in existence for nearly 50 years (in current and previous editions; AERA, APA & NCME, 1985; 1999), and have been relied upon in numerous legal proceedings concerning testing in state and federal courts. (Note 6) One specific provision of these standards reads as follows:

Standard 13.7 In educational settings, a decision or characterization that will have a major impact on a student should not be made on the basis of a single test score. Other relevant information should be taken into account if it will enhance the overall validity of the decision.

. . . It is important that in addition to test scores, other relevant information (e.g., school record, classroom observation, parent report) is taken into account by the professionals responsible for making the decision. (AERA, APA & NCME, 1999, pp. 146-47) (Note 7)

It seems clear that the practice in Texas of controlling award of high school diplomas on the basis of TAAS test scores in isolation without weighing other relevant information such as students' grades in high school (HSGPA) is contrary to this provision of the 1999 Standards for Educational and Psychological Testing (and the corresponding
provision of the 1985 *Standards*).

Witnesses for the state of Texas during the TAAS trial (Susan Phillips and William Mehrens) disputed my interpretation of this standard. Here is how Judge Prado summarized and resolved the dispute in his decision:

There was little dispute at trial over whether this standard exists and applies to the TAAS exit-level examination. What was disputed was whether the TAAS test is actually the sole criterion for graduation. As the TEA points out, in addition to passing the TAAS test, Texas students must also pass each required course by 70 percent. See Texas Admin. Code § 74.26(c).

Graduation in Texas, in fact, hinges on three *separate and independent* criteria: the two objective criteria of attendance and success on the TAAS examination, and the arguably objective/subjective criterion of course success. However, as the Plaintiffs note, these factors are not weighed with and against each other; rather, failure to meet any single criterion results in failure to graduate. Thus, the failure to pass the exit-level exam does serve as a bar to graduation, and the exam is properly called a "high-stakes" test.

On the other hand, students are given at least eight opportunities to pass the examination prior to their scheduled graduation date. In this regard, a single TAAS score does not serve as the sole criterion for graduation. The TEA presented persuasive evidence that the number of testing opportunities severely limits the possibility of "false negative" results and actually increases the possibility of "false positives," a fact that arguably advantages all students whose scores hover near the borderline between passing and failing. (Prado 2000, pp. 14-15)

Nonetheless, I believe that my interpretation of this standard is more in keeping with preponderance of professional opinion than are the narrow interpretations offered by the witnesses for the state of Texas. This may be illustrated by reference to the 1999 report from the Board on Testing and Assessment of the Commission on Behavioral and Social Sciences of the National Research Council.

As a result of increasing controversy over high stakes testing, the U.S. Congress passed legislation in 1997 requesting that the National Academy of Sciences undertake a study and make recommendations regarding the appropriate use of tests for student grade promotion, tracking and graduation (Heubert & Hauser, 1999, p. 1). The resulting report *High Stakes: Testing for Tracking, Promotion, and Graduation* specifically cites Standard 8.12 of the 1985 joint standards and clearly points out that a compensatory or sliding scale approach to using test scores in combination with grades would be "more compatible with current professional standards" than using an absolute cut-off score on a test to control high school graduation (Heubert & Hauser, 1999, pp. 165-66). More generally, this National Research Council report recommends:

High stakes decisions such as tracking, promotion, and graduation should not automatically be made on the basis of a single test score but should be buttressed by other relevant information about students' knowledge and skills such as grades, teacher recommendations and extenuating circumstances. (Heubert & Hauser, 1999, p. 279) (Note 8)

Ironically enough, reliance on TAAS scores in isolation to control award of high school diplomas in Texas is even contrary to the following passage from the TEA's own *Texas Student Assessment Program Technical Digest*:
All test result uses regarding individual students or groups should incorporate as much data as possible. . . . Student test scores should also be used in conjunction with other performance indicators to assist in making placement decisions, such as whether a student should take a reading improvement course, be placed in a gifted and talented program or exit a bilingual program. (pp. 2-3)

In sum, the state of Texas's use of TAAS scores in isolation, without regard to students' high school grades, to control award of high school diplomas, is contrary not only to both professional standards regarding test use and the advice of the recent NRC report, but also to the TEA's own advice on the need to use test results in conjunction with other performance indicators.

4.3 Passing scores on TAAS Arbitrary and Discriminatory

The problem of using TAAS scores in isolation to control award of high school diplomas is exacerbated by the fact that the passing scores set for TAAS are arbitrary and discriminatory. This is important because when a pass or cut score is set on a test, the validity of the test depends not just on test content, administration and scoring, but also on the manner in which the passing score is set.

The 1999 Standards for Educational and Psychological Testing state:

Standard 4.19 When proposed score interpretations involve one or more cut scores, the rationale and procedures used for establishing cut scores should be clearly documented. (AERA, APA & NCME, 1999, p. 59)

Also, standard 2.14 says that "Where cut scores are specified for selection or classification, the standard errors of measurement should be reported in the vicinity of each cut score (AERA, APA & NCME, 1999, p. 35). (Note 9)

Considerable technical and professional literature has been published on alternative methods for setting passing scores on tests. Glass (1978) wrote an early critique of methods of setting passing scores that questioned the very advisability of even attempting to make this use of tests. In 1986, Ronald Berk published "A consumer's guide to setting performance standards on criterion-referenced tests" (Review of Educational Research, 56:1, 137-172) in which he reviewed 38 different methods for setting standards (or pass or cut-scores) on standardized tests. (Note 10)

I sought to learn exactly how the passing scores were set on the TAAS in 1990 and to obtain copies of any data that were used in the process of setting passing scores on the TAAS exit test. The most complete account of the process by which the passing scores were set is provided in Appendix 9 of the Texas Student Assessment Program Technical Digest for the Academic Year 1996-1997, (TEA, 1997, pp. 337-354). Specifically contained in this appendix are 1) a memo dated July 14, 1990, from Texas Education Commissioner Kirby to members of the state Board of Education (including a summary of results from a field test of the TAAS) and 2) Minutes of the State Board of Education meeting in July 1990 at which the passing scores on the grade 10 TAAS were established.

In his memo, Commissioner Kirby recommended a passing score of 70% correct for the exit level of TAAS, but also recommended that this standard be phased in over a period of three years, with the passing score of 60% proposed for Fall 1990. After considerable discussion, the State Board voted unanimously to adopt the
recommendations of the commissioner regarding the Texas Assessment of Academic Skills, specifically that: "For the Academic Skills Level, a minimum standard of 70% of the test items must be answered correctly."

Following a statement by a Dr. Crawford about the importance of giving "notice regarding the standard required for graduation from high school . . . to those students who will be taking the exit level test" (p. 6/353), the Board also voted 11 to 3 in favor of an amendment to the original proposal to "give notice that the 1991-92 standard will be 70" (p. 7/354).

What struck me about this record of how the passing score on the TAAS exit test was set are the following:

1. The process was not based on any of the professionally recognized methods for setting passing standards on tests;
2. It appears to have failed completely to take the standard error of measurement into account; and,
3. As I explain below, the process yielded a passing score that effectively maximized the adverse impact of the TAAS exit test on Black and Hispanic students.

Before I elaborate on the latter point, let me emphasize that from the available record I have done my utmost to understand the rationale that motivated the Board to set the passing score where it did, namely at 70% correct. As best I can tell from the record, the main reasons for setting the passing score at 70% correct appear to have been that this is where the passing score had been set on TEAMS and this level was suggested by the Texas Education Code. The minutes of the Board meeting report that "the Commissioner cited the portion of the Texas Education code that requires 70 percent as passing (Attachment A), explaining that there is a rationale for aiming at 70 percent of test items as the mastery standard" (p. 1/348).

In my view this is simply not a reasonable or professionally sound basis for setting a passing standard on an important test such as the TAAS exit test. Indeed from the available record it is not even clear that the Texas code cited by the Commissioner was actually referring to anything more than the passing standard for course grades. Moreover, the minutes to the July 12, 1990, meeting also report the following remarks by Dr. Crawford: "Testing is driving a curricular program, which means that the curriculum is not at the place where you want it to be when you start out." She commented that "70 only has whatever value that is given to it, and in testing 70 is not the automatic passing standard on every test" (p. 4/351).

In sum, the process used in setting the passing scores on the TAAS exit test in 1990 did not adhere to prevailing professional standards regarding the setting of passing scores on standardized tests. For example, from the record available, it is clear that the process used to set the passing score on the TAAS exit test in 1990 failed to meet all six criteria of "technical adequacy" described in Berk's (1986) review of criteria for setting performance standards on criterion-referenced tests—a review published in a prominent education research journal, and of which TEA officials surely should have been aware in 1990.

**TAAS cut score study.** To understand more fully the process by which the TAAS passing scores were set in 1990, I requested a copy of the TAAS field test data that were presented to the Board of Education in the meeting at which it set the passing score on the TAAS-X. Using these data, I undertook a study (with the assistance of Boston College doctoral student Cathy Horn) which came to be called our "TAAS cut score study." In this study, we asked individuals, reviewing the data available to the Texas
Board of Education in July 1990 to select the passing scores (or cut scores) students would be need to attain in order to pass the TAAS reading and math tests. For both the reading and math tests, each research subject was presented with a graph showing the percentage of students, separately for White, Hispanic and Black ethnic groups, passing each number of percent correct answers on the field or pilot test of the TAAS exit test in 1990. These graphs are shown in Figures 4.1 and 4.2 below.

Each person in the cut score study was then presented with the following instructions:

The following graph presents the percentage of students passing the reading / math section of the Texas Assessment of Academic Skills (TAAS) at each number of questions answered correctly. Choose the number of questions
correct that most clearly differentiates White students (represented by a black line) from Black and Hispanic students.

Respondents could then ask clarifying questions before selecting a response. After a pilot test of the cut score study in 1998, Ms. Horn (a native of Texas and secondary school teacher there before she came to Boston College for graduate studies) extended the cut score study to nine Texans. The exercise was administered, by phone or in person, to 9 individuals residing in the state of Texas. (Those individuals who were interviewed by phone had paper copies of the Figure 4.1 and 4.2 graphs and the prompt for the exercise in front of them when they selected cut points.) The professions of the nine respondents are listed below.

Respondents (all currently living in Texas):
2 teachers
3 engineers
2 college students
1 financial analyst
1 director of communications

The cut or passing scores selected by these nine individuals as most clearly differentiating between White students and Black/Hispanic students are shown in Table 4.3 below.

**Table 4.3**

<table>
<thead>
<tr>
<th>Person</th>
<th>Reading</th>
<th>Math</th>
</tr>
</thead>
<tbody>
<tr>
<td>Person 1</td>
<td>34</td>
<td>34</td>
</tr>
<tr>
<td>Person 2</td>
<td>35</td>
<td>37</td>
</tr>
<tr>
<td>Person 3</td>
<td>35</td>
<td>38</td>
</tr>
<tr>
<td>Person 4</td>
<td>34</td>
<td>37</td>
</tr>
<tr>
<td>Person 5</td>
<td>36</td>
<td>40</td>
</tr>
<tr>
<td>Person 6</td>
<td>33</td>
<td>40</td>
</tr>
<tr>
<td>Person 7</td>
<td>34</td>
<td>37</td>
</tr>
<tr>
<td>Person 8</td>
<td>36</td>
<td>43</td>
</tr>
<tr>
<td>Person 9</td>
<td>44</td>
<td>44</td>
</tr>
</tbody>
</table>

**Summary**

<table>
<thead>
<tr>
<th></th>
<th>Reading</th>
<th>Math</th>
</tr>
</thead>
<tbody>
<tr>
<td>Minimum</td>
<td>33</td>
<td>34</td>
</tr>
<tr>
<td>Maximum</td>
<td>44</td>
<td>44</td>
</tr>
<tr>
<td>Mean</td>
<td>35.7</td>
<td>38.9</td>
</tr>
<tr>
<td>Median</td>
<td>35</td>
<td>38</td>
</tr>
</tbody>
</table>
As shown, respondents selected passing scores ranging from 33 to 44 on the reading test and from 34 to 44 for the math test. The median value across all nine respondents was 35 for the reading test and 38 for the math test.

The passing scores of 70% correct for the TAAS exit test recommended by Commissioner Kirby and accepted by the Board of Education in July 1990 were 34 for the reading test and 42 for the math test. The results of our cut score study show that if the intent in setting passing scores based on the TAAS field test results in July 1990 had been discriminatory, i.e., to set the passing scores so that they would most clearly differentiate between White students and Black/Hispanic students, then the passing scores would have been set just about where the Board of Education did in fact set them.

At the same time, there is no evidence of which I know, in the record of the process of setting passing scores on the TAAS in 1990, that the explicit intent of either Commissioner Kirby or the Board was discriminatory. However, the available record shows no indication that Commissioner Kirby, the TEA or the Board relied on any professionally recognized method for setting passing scores on the test, and the passing scores set were indeed consistent with those that would have been set, based on the TAAS field test results, if the intent had been discriminatory.

Use of measurement error in setting passing scores. The reason the setting of passing scores on a high stakes test such as the TAAS is so important is that the passing score divides a continuum of scores into just two categories, pass and fail. Doing so is hazardous because all standardized test scores contain some degree of measurement error. Hence, the 1985 Standards for Educational and Psychological Testing and other professional literature clearly indicate the importance of considering measurement error and consequent classification errors in the process of setting passing scores on tests.

Before discussing this topic further, two introductory explanations may be helpful. First, from the available record of the July 1990 meeting of the Board of Education, there is no indication that consideration of measurement error entered into the Board's deliberations. Second, the issue of measurement and classification errors regarding TAAS was addressed, as far as I know at least in the 1993-94 and 1996-97 editions of Texas Student Assessment Program Technical Digest. Unfortunately there are two serious errors in the manner in which these issues are addressed. Before explaining the nature of these errors, let me first summarize what the 1996-97 edition of Texas Student Assessment Program Technical Digest says about test reliability, standard error of measurement and classification errors.

Chapter 8 of the 1996-97 Technical Digest, entitled "reliability" provides a brief discussion of internal consistency estimates and formulas for calculating internal consistency reliability estimates (p.41). This is followed (p. 42) by a discussion of (and formulas for) calculating standard errors of measurement from reliability estimates. These discussions provide references to appendix 7 which shows data to indicate that for the Spring 1997 administration of TAAS at grade 10 (administered to 214,000 students) the internal consistency estimates for the TAAS math, reading and writing sub-tests were 0.934, 0.878 and 0.838, respectively; and the corresponding standard errors of measurement were 2.876, 2.352 and 2.195.

This represents the first serious error in the technical report's handling of measurement and classification error. Specifically, while the technical report bases the calculation of standard error of measurement on internal consistency reliability estimates, it clearly should have been based on test-retest or alternate-forms reliability estimates. Test-retest reliability refers to the consistency of scores on two administrations of a test. Alternate-forms reliability refers to the consistency of scores on two different forms or
versions of the same test. Since the purpose of TAAS testing is not simply to estimate students' performance on one version of the TAAS test, but to estimate their competence in reading, math and writing, in general, as might be measured by any version of the relevant TAAS tests, alternate-forms reliability is more appropriate for assessing reliability than is internal consistency reliability. As Thorndike and Hagen (1977, p. 79) point out in their textbook on measurement and evaluation, "evidence based on equivalent test forms should usually be given the most weight in evaluating the reliability of a test."

In general, alternate forms test reliability tends to be lower than internal consistency reliability. Hence, it seems clear to me that the figures reported in the 1996-97 Technical Digest overestimate the relevant reliability of grade 10 TAAS test scores and underestimate the standard error of measurement associated with TAAS scores.

I have attempted to estimate the alternate-forms reliability of TAAS test scores using two independent sources of data. First I employed the cross-tabulations reported by Linton & Debeic (1992) of test-retest data on students in several large Texas districts who took the TAAS exit level test in October 1990 and again in April 1991. Using the Linton & Debeic cross tabular results, I calculated the following test-retest correlations: TAAS-Reading 0.536; TAAS-Math 0.643; and TAAS-Writing 0.555. Second, as part of the background work for the TAAS case, Mark Fassold developed a remarkable longitudinal database of all 1995 sophomore students in Texas and their TAAS scores on up to ten different administrations of TAAS:

1  March 1995
2  May 1995
3  July 1995
4  October 1995
5  March 1996
6  May 1996
7  July 1996
8  October 1996
9  February 1996
10 April 1996

At my request Mr. Fassold ran an analysis of all test-retest correlations on this cohort of students (total N of about 230,000). Correlations were calculated separately by ethnic group and for TAAS Reading and Math tests. Given 16 different test-retest possibilities this yielded 214 different coefficients (2 x 16 x 6 ethnic groups). Results varied widely (in part because in some comparisons sample sizes were very small). Overall, however, the observed test-retest correlations tended to cluster in the 0.30 to 0.50 range.

These test-retest correlations based on both the Linton-Debeic and Fassold data are, however, attenuated in that in both data sets only students who failed a TAAS test took it again. There are methods for correcting observed test-retest correlations for such attenuation (see Haney, Fowler and Wheelock, 1999, for an example), but as a more conservative approach here, let me simply discuss what previously published literature suggests about the relationships between test-retest and internal consistency reliability.

As mentioned above, the 1996-97 Technical Digest cites internal consistency reliability estimates for the three grade 10 TAAS sub-tests of 0.934, 0.878 and 0.838, and standard errors of measurement of 2.876, 2.352 and 2.195. It is common for tests which
show internal consistency reliability of about 0.90 to show alternate forms reliability of 0.85 or 0.80 (see for example, Thorndike & Hagen, 1977, p. 92). On page 42 of the 1996-97 Technical Digest, the example is shown in which a test with an internal consistency reliability of 0.90 (and a standard deviation of 6.3) is estimated to have a standard error of measurement of 2.0. However, if instead of an internal consistency reliability of 0.90, we were to use in these calculations an alternate forms reliability of 0.85 or 0.80, the resulting standard errors of measurement would be 2.44 and 2.82. This suggests that the appropriate standard errors of measurement for the TAAS tests may be on the order of 20 to 40% greater than the estimates reported in the TAAS 1996-97 Technical Digest.

The second serious error in the technical report's handling of measurement and classification error occurs on pages 30 and 31 in a section labeled "Exit level testing standards and the standard error of measurement." Here the authors of the 1996-97 Technical Digest point out that a student with a "true achievement level at the passing standard would be likely to pass on the first attempt only 50% of the time" (p. 31). This passage then goes on to assert that "if such a student has attempted that test eight times, the student's passing is almost assured (probability of passing is 99.6%)" (p. 31). In other words, the chances of a minimally qualified student failing the TAAS eight times and being misclassified as not having the requisite skills is only 0.4% (0.50 to the 8th power is 0.0039).

This calculation strikes me as erroneous, or at least potentially badly misleading, because the authors have presented absolutely no evidence to show the probability that a student who fails the TAAS will continue to take the test seven more times. As I explain later, available evidence suggests that students who fail the TAAS grade 10 test more than once or twice are likely to be held back in grade and to drop out of school long before they reach grade 12 by which time they would have had a chance to take the TAAS exit test eight times. Since 0.50 to the second power is 0.25; and to the third power is 0.125, this indicates that a student with a "true achievement level at the passing standard" who takes the TAAS twice or three times, before becoming discouraged and not taking the test again, has a 25% or 12.5% chance of being misclassified as failing.

Before proceeding to present evidence bearing on this point, let me discuss how the standard error of measurement might usefully have been taken into account in adjusting passing scores. Because of the error of measurement in test scores, when scores are used to make pass-fail decisions about students, two kinds of classification errors can occur. A truly unqualified student can pass the test (a false pass) or a truly qualified student can fail the test (a false failure). How one thinks about the balance of these two misclassification errors depends on the risks (or benefits and costs) associated with each type of misclassification. If one were confident that a student failing TAAS would receive special attention and support educationally, one might be inclined to weigh false passes as more serious than false failures. If on the other hand, one thought that students failing TAAS were unlikely to receive effective instruction, and instead merely to be retained in grade 10 and to be stigmatized as failures, then one would probably feel that false failures would be more harmful than false passes.

Here is how Berk (1986) discussed this point:

Assessing the relative seriousness of these consequences, is a judgmental process. It is possible to assign plusses (benefits) and minuses (costs or losses) to the consequences so that the cutoff scores can be set in favor of a specific error reduction rate. A loss ratio (benefits: losses) can be specified for each decision application with the cutoff score adjusted accordingly.
To study the relative risks associated with the two kinds of classification errors associated with a high school graduation test, with the assistance of Kelly Shasby, (a doctoral student in the Educational Research, Measurement and Evaluation program at Boston College), I undertook what came to be known as our "risk analysis" study.

The survey form used in the risk analysis study was entitled "Survey of risk associated with classification decisions" and opened with the following introduction:

When classifying large numbers of individuals using standardized exams, two different kinds of mistakes are made. Some people will be falsely classified as "qualified" or "passing" while others will be falsely classified as "unqualified" or "failing." There is a degree of risk associated with mistakes of this kind, both for the individual who is incorrectly classified and for the society in which that individual lives. We would like your help in assessing the severity of the risk, or possible harm, caused to individuals and to society when mistakes are made on a number of different types of standardized tests.

The purpose of this survey is to assess the public's perception of misclassifications of individuals. These misclassifications can have an impact on the individual and on the society in which that individual lives. This impact has the potential to be harmful, and we are interested in determining how harmful the public thinks different misclassifications can be.

On a scale from 1 to 10, 1 being "minimum harm" and 10 being "maximum harm," rate each scenario with respect to the degree of harm it would cause that individual and then the degree of harm it would cause society. Then circle the number, which corresponds, to the rating you chose.

After this introduction, respondents were asked to rate the risk on a 1 to 10 scale of harm associated with 16 different misclassifications that might result from classifying people pass-fail based on standardized test results. Respondents were asked to rate separately the harm to individuals and to society—and to give credit where it is due, this distinction, a clear improvement over the initial version of our survey, was suggested by Ms. Shasby. Specifically, survey respondents were asked to rate the degree of harm, separately for individuals and society, associated with the following kinds of misclassification:

1. A kindergartner who is ready to enter school is denied entrance.
2. A kindergartner who is not ready to enter school is granted entrance.
3. An airline pilot who is not qualified is given a license to fly.
4. An airline pilot who is qualified is denied a license to fly.
5. A qualified high school student is denied a diploma.
6. An unqualified high school student is granted a diploma.
7. A qualified accountant is denied certification.
8. An unqualified accountant is granted certification.
9. A qualified student is denied promotion from grade eight to grade nine.
10. An unqualified student is granted promotion from grade eight to grade nine.
11. A qualified doctor is denied a license to practice.
12. An unqualified doctor is granted a license to practice.
13. A qualified candidate is denied admission into college.
14. An unqualified candidate is granted admission into college.  
15. A qualified teacher is denied certification.  
16. An unqualified teacher is granted certification.

The risk survey form was sent to a random sample of 500 secondary teachers in Texas (specifically only math and English/Language Arts teachers) on May 23, 1999. As of June 30, 1999, we had received 66 responses (representing a response rate of 13.2%). (Note 11)  
Table 4.4 below summarizes the results of the risk analysis survey.

**Table 4.4**  
**Results of Risk Analysis Survey with Secondary Teachers in Texas**

<table>
<thead>
<tr>
<th>Event Description</th>
<th>For individual</th>
<th>For society</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. A kindergartner who is ready to enter school is denied entrance.</td>
<td>6.45 ± 2.67</td>
<td>3.94 ± 2.64</td>
</tr>
<tr>
<td>2. A kindergartner who is not ready to enter school is granted entrance.</td>
<td>7.20 ± 2.23</td>
<td>5.06 ± 2.71</td>
</tr>
<tr>
<td>3. An airline pilot who is not qualified is given a license to fly.</td>
<td>8.36 ± 2.32</td>
<td>9.55 ± 1.00</td>
</tr>
<tr>
<td>4. An airline pilot who is qualified is denied a license to fly.</td>
<td>7.74 ± 2.37</td>
<td>4.39 ± 2.99</td>
</tr>
<tr>
<td>5. A qualified high school student is denied a diploma.</td>
<td>9.11 ± 1.69</td>
<td>6.39 ± 2.58</td>
</tr>
<tr>
<td>6. An unqualified high school student is granted a diploma.</td>
<td>6.85 ± 2.72</td>
<td>7.74 ± 2.26</td>
</tr>
<tr>
<td>7. A qualified accountant is denied certification.</td>
<td>8.65 ± 1.50</td>
<td>5.32 ± 2.62</td>
</tr>
<tr>
<td>8. An unqualified accountant is granted certification.</td>
<td>8.65 ± 1.50</td>
<td>5.32 ± 2.62</td>
</tr>
<tr>
<td>9. A qualified student is denied promotion from grade eight to grade nine.</td>
<td>8.89 ± 1.52</td>
<td>6.15 ± 2.39</td>
</tr>
<tr>
<td>10. An unqualified student is granted promotion from grade eight to grade nine.</td>
<td>8.15 ± 2.01</td>
<td>7.80 ± 2.12</td>
</tr>
<tr>
<td>11. A qualified doctor is denied a license to practice.</td>
<td>8.80 ± 1.68</td>
<td>7.32 ± 2.64</td>
</tr>
<tr>
<td>12. An unqualified doctor is granted a license to practice.</td>
<td>7.15 ± 2.87</td>
<td>9.37 ± 1.72</td>
</tr>
<tr>
<td>13. A qualified candidate is denied admission into college.</td>
<td>8.83 ± 1.73</td>
<td>6.30 ± 2.43</td>
</tr>
<tr>
<td>14. An unqualified candidate is granted admission into college.</td>
<td>6.08 ± 2.66</td>
<td>6.08 ± 2.66</td>
</tr>
<tr>
<td>15. A qualified teacher is denied certification.</td>
<td>8.64 ± 1.76</td>
<td>8.38 ± 2.13</td>
</tr>
<tr>
<td>16. An unqualified teacher is granted certification.</td>
<td>6.62 ± 2.84</td>
<td>9.15 ± 1.60</td>
</tr>
</tbody>
</table>

As this table shows, the risk associated with denying a high school diploma to a qualified student is for individuals the most severe risk associated with any of the
misclassification scenarios we asked respondents to rate. The only scenarios showing higher average risks are the risks for society associated with licensing an unqualified pilot (mean = 9.55), licensing an unqualified doctor (9.37) and licensing an unqualified teacher (9.15).

Particularly germane to our discussion of the setting of passing scores on the TAAS graduation test are the relative risks associated with denying a diploma to a qualified high school student (mean = 9.11) and granting a diploma to an unqualified student (6.85). These results indicate that the risk of denying a diploma to a qualified student is much more severe than granting a diploma to an unqualified student (the difference, by the way, is statistically significant).

These results indicate that if a rational passing score had been established on the TAAS exit test, the passing or cutoff scores should be adjusted downward in order to minimize overall risk. A common practice in setting passing scores on important tests is to reduce an empirically established passing score by one or two standard errors of measurement. While I want to stress that the passing scores of 70% correct on the TAAS are arbitrary, unjustified and discriminatory, we can see from Figures 4.1 and 4.2 what the consequences would be for Black and Hispanic pass rates (on the TAAS field test) if the passing scores of 70% had been corrected for error of measurement. Recall that the passing scores set by the Board on the field test administration of the TAAS were 34 items correct on the reading test and 42 on the math test. Recall also that the standard errors for the reading and math tests reported in the Technical Digest were in the range of 2.5 to 3.0 raw score points. Suppose that to take error of measurement into account, the initially selected passing scores of 34 and 42 were lowered 5 points, to 29 and 37 on the reading and math tests, respectively. What can be easily seen from Figures 4.1 and 4.2 is that these adjustments would have increased the passing rates for Black and Hispanic students about 12% on the math test and 20% on the reading test.

The foregoing results were presented in a written report before the TAAS trial (Haney, 1999) and also discussed during testimony at trial. Judge Prado (2000) apparently did not find these points persuasive for he commented merely that in setting the passing score on the TAAS tests, "the State Board of Education looked at the passing standard for the TEAMS test, which was also 70 percent, and also considered input from educator committees" (p. 11). Regarding the disparate impact of the passing score, he commented simply, "The TEA understood the consequences of setting the cut score at 70 percent" (p. 11).

4.4 Doubtful Validity of TAAS Scores

The Technical Digest on TAAS (TEA, 1997) contains an extremely short section (pp. 45-47) discussing test validity. Though this three-page passage mentions content, construct and criterion-related validity, it maintains that "the primary evidence for the validity of the TAAS and end-of-course tests lies in the content validity of the test" (TEA, 1997, p. 47). This discussion, it seems to me is woefully inadequate because test validation should never rest primarily on test content. Test validation refers to the interpretation and meaning of test scores and these depend not just on test content, but also on a host of other factors, such as the conditions under which tests are administered, and how results are scored and interpreted (e.g., in terms of a passing score, as discussed in the previous section).

Nonetheless, the TEA has previously undertaken a number of studies examining the relationship between TAAS scores and course grades. In one study, for example, it was reported that in one large urban district, 50% of the students who had received a
grade of B in their math courses failed the TAAS math test (TEA, 1996 Comprehensive Report on Texas Public Schools, pp. 14-15). Another summary finding was that when "TEA correlated exit level students' TAAS mathematics scores with the same students' course grades for several different mathematics courses in the 1992-93 school year . . . the correlation between TAAS scale scores and students' end-of-year grades was only moderately positive (0.32) . . . " (TEA, 1997, Technical Digest, p. 47). Inasmuch as this correlation is remarkably low in light of previous research that has generally shown test scores to correlate with high school grades in the range of 0.45 to 0.60 (see Haney, 1993, p. 58), as part of work on the TAAS case I sought to acquire the actual data set on which this TEA finding was based.

The data set in question contains records for 3,281 students in three districts that TEA documentation describes as "large urban district," "mid-sized suburban district," and "small rural district." The TEA has previously reported analyses of these data in "Section V: A study of correlation of course grades with Exit Level TAAS Reading and Writing Tests" pp. 189-197 in Student Performance Results 1994-95, Texas Student Assessment Program, TAAS and End-of-Course Examinations and Other Studies (Texas Education Agency, Austin, Texas, ND, but presumably 1995).

After opening the file and verifying its structure, I sought to confirm that the results reported by the TEA could be replicated. This was impossible to do precisely because TEA did not report results with great precision. Nonetheless, initial results corresponded reasonably well with what TEA reported. Also, it should be noted that while the data file included records on a number of grade 11 students, I restricted most analyses to grade 10 students pooled across the three districts, though the bulk of this sample (> 2,400 cases out of 3,300) comes from the one large urban district. Then we calculated basic descriptive statistics on variables of interest, in particular scores for the TAAS reading and writing test administered in March 1995 and grades for the English II courses completed in May 1995 (these data were provided by the districts to the Student Assessment Division of TEA.) Next we calculated relationships between variables. Table 4.5 shows the intercorrelations between the three TAAS test scores (writing, reading and math) and English II course grades. Given the size of this sample (>3,000) all of these correlation coefficients are statistically significant at the 0.01 level.

<table>
<thead>
<tr>
<th></th>
<th>Write SS</th>
<th>Read SS</th>
<th>Math SS</th>
<th>Grade</th>
</tr>
</thead>
<tbody>
<tr>
<td>Write SS</td>
<td>1.00</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Read SS</td>
<td>0.50</td>
<td>1.00</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Math SS</td>
<td>0.51</td>
<td>0.69</td>
<td>1.00</td>
<td></td>
</tr>
<tr>
<td>Grade</td>
<td>0.32</td>
<td>0.34</td>
<td>0.37</td>
<td>1.00</td>
</tr>
</tbody>
</table>

Note the magnitudes of the correlations between English II course grades and TAAS scores. They are all in the range of 0.32 to 0.37. As indicated above, previous studies have generally shown test scores to correlate with high school grades in the range of 0.45 to 0.60. Contrary to expectations, English II grades correlate more highly with TAAS math scores (0.37) than with writing (0.32) or reading (0.34) scores. Note also the odd intercorrelations among TAAS scores. The TAAS math scores correlate at the level
of 0.69 with the TAAS reading scores, while the TAAS reading scores correlate at the level of 0.50 with the TAAS writing scores. This is contrary to the expectation that scores of two verbal measures (of reading and writing) should correlate more highly with one another than with a measure of quantitative skills. These results cast doubt on the validity and the reliability of TAAS scores.

People unfamiliar with social science research doubtless find it hard to make sense of correlation coefficients in the range of 0.32 to 0.37. Hence to provide a visual representation, Figure 4.3 shows a scatterplot of the relationship between TAAS reading scores and English II grades. As can be seen from this figure, the relationship between these two variables is a quite weak. Students with grades in the 70 to 100 range have TAAS reading scores from well below 40 to well over 80. Conversely, students with TAAS reading scores in the 80 to 100 range have English II grades from well below 40 to well over 80.

![Figure 4.3 Scatterplot of TAAS Reading Scores and English II Grades](image)

I next examined whether there were differences in the relationships between TAAS scores and English II grades across ethnic groups. Table 4.6 provides an example of the relationship between passing and failing TAAS and passing or failing in terms of English II course grades for Hispanics, Blacks and Whites. As can be seen from this table, of those students who passed their English II courses in the spring of 1995, 27-29% of Black and Hispanic students failed the TAAS reading test taken the same semester as their English courses compared with 10% of White students. In other words, of grade 10 students in these three districts who are passing their English II courses, the rate of failure on the TAAS reading test for Black and Hispanic students is close to triple that of White students. A similar, but slightly smaller, disparity is apparent on the TAAS writing sub-test.

Table 4.6
Rates of Passing and Failing TAAS and English II Course

<table>
<thead>
<tr>
<th>TAAS-Exit Test Results</th>
<th>Black students</th>
<th>Hispanic students</th>
<th>White students</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Reading</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>English II Course</strong></td>
<td>Failed</td>
<td>Passed</td>
<td>Failed</td>
</tr>
<tr>
<td>Failed N</td>
<td>39</td>
<td>23</td>
<td>242</td>
</tr>
<tr>
<td>(%)</td>
<td>10.1%</td>
<td>5.9%</td>
<td>11.0%</td>
</tr>
<tr>
<td>Passed N</td>
<td>111</td>
<td>214</td>
<td>596</td>
</tr>
<tr>
<td>(%)</td>
<td>28.7%</td>
<td>55.3%</td>
<td>27.0%</td>
</tr>
<tr>
<td><strong>Writing</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>English II Course</strong></td>
<td>Failed</td>
<td>Passed</td>
<td>Failed</td>
</tr>
<tr>
<td>Failed N</td>
<td>33</td>
<td>29</td>
<td>173</td>
</tr>
<tr>
<td>(%)</td>
<td>8.5%</td>
<td>7.5%</td>
<td>7.8%</td>
</tr>
<tr>
<td>Passed N</td>
<td>69</td>
<td>256</td>
<td>366</td>
</tr>
<tr>
<td>(%)</td>
<td>17.8%</td>
<td>66.1%</td>
<td>16.6%</td>
</tr>
</tbody>
</table>

Such a disparity can result from several causes. First, if the TAAS reading test is in fact a valid and unbiased test of reading skills, the fact that close to 30% of Black and Hispanic students who are passing their sophomore English courses failed the TAAS reading test, as compared with only 10% of White students must indicate that minority students in these three districts are simply not receiving the same quality of education as their White counterparts—especially when one realizes, as I will show in Part 5 of this article that by 1995 Black and Hispanic students in Texas statewide were being retained in grade 9 at much higher rates than White students. The only other explanation for the sharp disparity is that the TAAS tests and the manner in which they are being used (with a passing score of 70% correct) are simply less valid and fair measures of what Black and Hispanic students have had an opportunity to learn, as compared with White students.

These analyses were reported in the July 1999 report (Haney, 1999) and discussed in direct testimony and cross-examination during the TAAS trial in September 1999. Here is how Judge Prado interpreted these findings in his January 7 ruling:

The Plaintiffs provided evidence that, in many cases, success or failure in relevant subject-matter classes does not predict success or failure in that same area on the TAAS test. See Supplemental Report of Dr. Walter Haney, Plaintiff's expert, at 29-32. In other words, a student may perform reasonably well in a ninth-grade English class, for example, and still fail the English portion of the exit-level TAAS exam. The evidence suggests that the disparities are sharper for ethnic minorities. Id. at 33. However, the TEA has argued that a student's classroom grade cannot be equated to TAAS performance, as grades can measure a variety of factors, ranging from effort and improvement to objective mastery. The TAAS test is a solely objective
measurement of mastery. The Court finds that, based on the evidence presented at trial, the test accomplishes what it sets out to accomplish, which is to provide an objective assessment of whether students have mastered a discrete set of skills and knowledge. (Prado, 2000, p. 24)

With due respect to Judge Prado, I believe there are two flaws in this reasoning. First, Judge Prado interprets the disparities in the rates at which, among students who pass their English II courses, minorities fail the "English portion" of TAAS far more frequently than White students, as evidence of the need for "objective assessment" of student skills. Though he did not explicitly say so, his reasoning seems to be that an objective test is necessary because the grades of minority students are inflated. This interpretation, however, takes one specific finding out of the context in which I presented it, both in the Supplementary report (Haney, 1999, pp. 29-33) and in testimony at trial. In both cases, and as described above, it was shown that even if one ignores the question of possibly inflated grades, the intercorrelations among TAAS scores themselves (i.e., that reading and math scores correlate more highly than reading and writing scores) raise serious doubts about their validity.

Second, even if we assume the validity of TAAS tests and accept Judge Prado's reasoning that the lack of correspondence between English grades and TAAS reading and writing scores demonstrates the need for objective assessment of student mastery, the fact that "the disparities are sharper for ethnic minorities," represents prima facie evidence of inequality in opportunity to learn. Even if Black and Hispanic students' teachers are covering the same academic content as White students' teachers, that 27-29% of Black and Hispanic students who passed their English II course failed the TAAS reading test (as compared with 10% of White students) obviously must indicate that their teachers are not holding them to the same academic standards as the teachers of White students.

4.5 More appropriate use possible

This discussion leads naturally to a simple solution for avoiding reliance on test scores in isolation to make high stakes decisions about students. As previously mentioned, the recent High Stakes report of the National Research Council (Heubert & Hauser, 1999) states clearly that using a sliding scale or compensatory model combining test scores and grades would be "more compatible with current professional testing standards" than relying on a single arbitrary passing score on a test (Heubert & Hauser, 1999, pp. 165-66). Moreover this is exactly how test scores are typically used in informing college admissions decisions, such that students with higher high school grade point averages (GPA) need lower test scores to be eligible for admission, and conversely students with lower GPA need higher test scores. Ironically enough this is indeed exactly how institutions of higher education in Texas use admissions test scores in combination with GPA. For example, in 1998, the University of Houston required that in order to be eligible for admissions, high school students who had a grade point average of 3.15 or better needed to have SATI total scores of at least 820, but if their high school GPA was only 2.50, they needed to have SATI total scores of 1080 (University of Houston, 1998).

Literally decades of research on the validity of college admissions test scores show that such an approach, using test scores and grades in sliding scale combination produces more valid results than relying on either GPA or admissions test scores alone (Linn, 1982; Willingham, Lewis, Morgan & Ramist, 1990). Moreover, such a sliding scale approach generally has been shown to have less disparate impact on ethnic minorities (and women) than relying on test scores alone (Haney, 1993).
The tendency for a sliding scale approach to have smaller adverse impact on minorities can be illustrated with the data on TAAS scores and English II grades discussed in the last section. Texas now effectively uses a double-cut or conjunctive model of decision-making, whereby students currently must have a grade of 70 in their academic courses (such as English II) and a score of 70 on TAAS to graduate from high school. These requirements are illustrated in Figure 4.4 (which is the same as Figure 4.3 except that a vertical line has been added to represent the 70-grade requirement and a horizontal line has been added to represent the TAAS 70-score requirement.

Note also that the data shown in Figure 4.4 are the same as those summarized in the top portion of Table 4.6. As indicated there, 80.4% of white students in this sample passed both the English II course and the TAAS reading test, while only 10.1% of White students passed English II and failed the TAAS reading test. In contrast, 53-55% of Black and Hispanic students passed both the course and the test, but 27-29% of Black and Hispanic students passed English II, but failed the TAAS test.

Suppose now that instead of applying a double cut rule so that students have to have scores of 70 in both the course and the test to pass, they need to have a minimum of 140 combined. This circumstance is illustrated in Figure 4.5, below.
As can be seen, under such a sliding scale approach, higher grades can compensate for lower test scores and vice versa (that is why the sliding scale approach is sometimes called a compensatory model). Under this approach, the number of Black and Hispanic students passing would increase from 1,395 to 1,765—a 27% increase. Under a sliding scale approach, the number of White students passing would also increase slightly (from 436 to 487), but since the latter increase is smaller proportionately, the disparate impact on Black and Hispanic students would be reduced.

The sliding scale decision rule illustrated here (TAAS-R + Eng II grade > 140) was chosen merely for illustrative purposes. As with college admissions tests, in practice such a sliding scale approach ought to be based on empirical validation studies. But the example illustrates the way in which an approach more in accord with professional standards would significantly reduce adverse impact. The literature on college admissions testing strongly suggests it would yield more valid decisions too.
5. Missing Students and Other Mirages

As previously mentioned, dropout rate is one of the indicators used in the TEA accountability system for rating Texas districts and campuses. Also, as summarized in Section 3.3 above, the TEA has reported that dropout rates have been decreasing in Texas during the 1990s. However, in 1998 when I began studying what had been happening in Texas schools, I quickly became suspicious of the validity of the TEA-reported dropout data. At least one independent organization in Texas had previously challenged TEA's "dropout calculation methodology" (TRA, 1998, p. 2). Moreover, two independent sources were reporting substantially higher rates of dropouts (or attrition) or, conversely, lower rates of high school completion than would be implied by TEA dropout data (Fassold, 1996; IDRA, 1996).

Hence, to examine independent evidence on recent patterns of high school completion in Texas and possible effects of the TAAS on grade enrollment patterns and high school completion, I assembled data on the numbers of White, Hispanic and Black students enrolled in every grade (kindergarten to grade 12) in Texas over the last two decades. (Note 12)

Before describing analyses of these data, three additional points should be made. First, in assembling this data set, we have taken care to double-check the accuracy of all data input (in this context, "we" refers to myself and Damtew Teferra, a Boston College doctoral student who helped me assemble the Texas enrollment data set). Second, to my original set of data on grade enrollment by ethnic group for each year between 1975-76 and 1998-99, I added data on the numbers of high school graduates each year (provided to me, again, thanks to the kind assistance of Dr. Rincon and Terry Hitchcock). Third, I should mention that data on enrollments and graduates for 1998-99 were not available until recently and hence were not considered in my previous reports or in the TAAS trial in the Fall of 1999. Finally, in case others might wish to verify results shown below, or conduct other analyses of Texas enrollments over the last quarter century, I make available via this publication, the set of data I have assembled (see, Appendix 7).

5.1 Progress from Grade 9 to High School Graduation

In this analysis, I simply took the numbers of White, Black and Hispanic Texas high school graduates by year and divided each of these numbers respectively by the number of White, Black and Hispanic students enrolled in grade nine three years earlier. The resulting ratios show the proportion of grade nine students for each ethnic group who progress on time to high school graduation three-and-a-half years later. The results of this analysis are shown in Figure 5.1.
Figure 5.1 shows that between 1978 and 1985-86, the ratio of HS graduates to grade nine students three years earlier ranged between 0.72 and 0.78 for White students and between 0.57 and 0.64 for Black and Hispanic students. Between 1985-86 and 1989-90 these ratios declined slightly for all ethnic groups, from 0.72 to 0.70 for Whites, from 0.59 to 0.57 for Blacks and from 0.57 to 0.56 for Hispanics. However, in 1990-91, the first year the TAAS high school graduation test was used, the ratios for all three groups evidence the most precipitous drops in the whole 20-year time series: for Whites from 0.699 to 0.640 (a drop of 0.059), for Blacks from 0.567 to 0.474 (a drop of 0.093) and for Hispanics from 0.564 to 0.476 (a drop of 0.088). In other words, the steep drop in this indicator of progress from grade 9 to high school graduation was about 50% greater for Black and Hispanic students than for White students.

In 1991, the ratios for all three ethnic groups showed a slight rebound, from 0.640 to 0.700 for Whites, from 0.474 to 0.518 for Blacks and 0.476 to 0.513 for Hispanics. In 1992-93, the first year in which the TAAS graduation requirement was fully implemented, Whites showed a minor decline, from 0.700 to 0.689, but for Blacks and Hispanics declines were larger: from 0.518 to 0.479 for Blacks and from 0.513 to 0.491 for Hispanics.

From full implementation of the TAAS as a requirement for high school graduation in Texas in 1992-93 (with the passing score set at 70%) until 1998-99, the ratio of HS graduates to grade nine students three years earlier has been just at or below 0.500 for Black and Hispanic students, while it has been just about 0.700 for White students.

Figure 5.2 presents another view of these data. This figure shows the ratio of the number of Texas high school graduates divided by the number of grade nine students three years earlier for White and Nonwhite students. What this figure shows even more clearly than the previous figure is that since the three-year period 1990-92 in which the TAAS exit test requirement was phased in, the gap in this ratio for White and Nonwhite students has widened substantially. Specifically, during the period 1978 through 1989, the average gap in the ratios graphed in Figure 5.2 was 0.146. However, the average gap
in the ratios for Whites and Nonwhites since the TAAS exit test requirement was fully implemented in 1992-93 has been 0.215. This indicates that the TAAS exit test has been associated with a 50% increase in the gap in progression from grade 9 to high school graduation for Nonwhite students as compared with White students.

5.2 Grade-to-Grade Progression Ratios

What happened between the late 1970s and the mid-1990's? (Note 13) Where did the decline in progression between grade nine and high school graduation occur for Black and Hispanic students? Was it at grade 10 when they first took the TAAS exit test, or in grade 12 after they had had a chance to take the TAAS-X as many as eight times?

To shed light on this question, I calculated the grade-to-grade progression ratios of the number of students enrolled in one grade divided by the number of students enrolled in the previous grade in the previous year, separately for the Black, Hispanic and White ethnic groups. Altogether, 858 such calculations were computed—13 grade transitions (from kindergarten to grade 1, etc., to grade 12 to high school graduation) for 22 years and three ethnic groups. Overall there was considerable consistency in these grade transition ratios. Across the last twenty years, and the 13 grade transitions, for the three ethnic groups, overall, transitions from one grade to another have been highly consistent, with 99 or 100% of each ethnic group, on average, progressing from one grade in one year to the next grade in the following year.

What the detailed results show, however, is that there are two sets of grade progression ratios that were highly unusual (greater than 1.24 or more than 2 standard deviations from the mean across all transition ratios; see Haney 1999, Table 5). First, in the decade 1976 to 1986, there were 25 grade progression ratios that exceeded 1.24. These were all for the grade 1/kindergarten ratios, and mostly for Black and Hispanic students, though there were a few years for which the comparable ratios for White students exceeded 1.24. It is likely that these high ratios resulted partly from a time when kindergarten attendance in Texas was not universal and many students entered school in grade 1 without previously having attended kindergarten.

Since 1990, there were more than a dozen grade progression ratios that exceeded 1.24. For each and every year from 1992-93 to 1998-99, the grade 9/grade 8 progression
ratio for Black and Hispanic students has exceeded 1.24, while the comparable ratio for White students has remained in the range of 1.08 to 1.11. As shown in Figure 5.3, since 1990 the grade 9/grade 8 progression ratio for Black and Hispanic students has risen dramatically, while the comparable rate for White students increased only slightly.

![Figure 5.3 TX Grade 8 to 9 Progression Ratios by Ethnic Group 1977 to 1999](image)

The data also reveal that before the mid-1980s, the grade 9/grade 8 progression ratios for Black and Hispanic students were only slightly higher than those for Whites. These results clearly indicate that since 1992 progress from grade 9 to high school graduation has been stymied for Black and Hispanic students not after grade 10 when they first take the TAAS exit test, but in grade nine before they take the test. These results clearly support the hypothesis advanced in my December 1998 report, namely that after 1990 schools in Texas have increasingly been retaining students, disproportionately Black and Hispanic students, in grade nine in order to make their grade 10 TAAS scores look better (Haney, 1998, pp. 17-18).

This hypothesis was discussed during the TAAS trial. In his ruling, Judge Prado held that "Expert Walter Haney's" hypothesis that schools are retaining students in ninth grade in order to inflate tenth-grade TAAS results was not supported with legally sufficient evidence demonstrating the link between retention and TAAS (Prado, 2000, p. 27). In Section 5.6 below, I present documentation that was not allowed into the TAAS trial as evidence to support the hypothesis. For now, however, suffice it to note that the pattern apparent in Figure 5.3 provides a clear explanation for one aspect of the Texas "miracle," namely, the apparent decrease in the racial gap in test scores (reviewed in Section 3.3 above). One clear cause for the decrease in the racial gap in grade 10 TAAS scores in the 1990s (see Table 3.2 and Figure 3.2) is that Black and Hispanic students are being increasingly retained in grade 9 before they take the grade 10 TAAS tests. Between 1989-90 (the year before TAAS was implemented) and the late 1990s, the
grade 9/grade 8 progression ratios for Black and Hispanic students grew from about 1.20 to 1.30, while the comparable ratio for White students remained at about 1.1.

It is apparent from Figure 5.3 that the higher rates of grade 9 retention of Black and Hispanic students, as compared White students, did not begin with TAAS. The results shown in Figure 5.3 indicate that the grade 9/grade 8 progression ratios for minorities began to diverge from those of White students in Texas in the 1980s, before TAAS and even before TEAMS. In an historical sense, then, TAAS and TEAMS could not have directly caused the steady increase since the early 1980s in the proportions of Black and Hispanics retained in grade 9. But the first statewide testing program in Texas, the TABS, did begin in 1980, just about the time the ratio of minority ninth graders to eighth graders began its upward climb, compared with the relative stability of this ratio for White students. Whatever the historical cause, the fact that by the end of the 1990s 25-30% of Black and Hispanic students, as compared with only 10% of White students, were being retained to repeat grade 9, instead of being promoted to grade 10, makes it clear that the apparent diminution in the grade 10 racial gap in TAAS pass rates is in some measure an illusion.

Data for the last two academic years, i.e., 1997-98 and 1998-99, provide a picture of how grade progression ratios compare across the grade levels. Specifically, Figure 5.4 shows the grade progression ratios for grades 1 through 12 and for graduates. For grades 1 through 12 these are simply the number of students enrolled in a particular grade in 1998-99 divided by the number enrolled in the previous grade in 1997-98. The only exception to this pattern is for graduates in which the ratio shown is the number of graduates in 1999 divided by the number enrolled in grade 12 in the fall of 1998-99.

As can be seen, for most grade levels the progression ratios are highly similar for Black, Hispanic and White students. Indeed for grades 2 through 8 all of the transition ratios are close to 1.00. Note however how sharply the transition ratios diverge for grades 9 and 10. In 1998-99, there were about 30% more Black and Hispanic students enrolled in grade 9 than had been enrolled in grade 8 in 1997-98 (as compared with about 10% more
Whites). Also, in 1998-99 there were 25-30% fewer Black and Hispanic students enrolled in grade 10 than had been enrolled in grade 9 in 1997-98. These data indicate that at the end of the 1990s even for students who had been going to school for virtually their entire careers under TAAS testing (a student in grade 9 in 1998-99, would have been in grade 1 in 1990-91, if not retained in grade), there remains a huge gap in progress in the early high school years for Black and Hispanic students as compared with Whites. As will be shown subsequently, after being retained to repeat grade 9 and/or 10, tens of thousands of students in Texas drop out of school.

5.3 Progress from Grade 6 to High School Graduation

The apparent increase in grade 9 retention rates suggests a need to revisit the question of rates of progress toward high school graduation. In Section 5.1 above, we saw that the rate of progress of Black and Hispanic students from grade 9 to high school graduation fell to about 50% after full implementation of the TAAS as a requirement for high school graduation in 1992-93. But now, having seen in section 5.2 that the rate of retention in grade 9 appears to have increased markedly for Black and Hispanic students in Texas during the 1990s, it is useful to revisit the question of rates of progress toward high school graduation using base years other than grade 9 as a starting point from which to chart progress. This is because the grade 9 to high school graduation progress ratio may be lowered because of the increasing numbers of students "bunching up" in grade 9.

A number of analyses have been conducted, examining the rates of progress from grades 6, 7, and 8 to high school graduation, six, five and four years later, respectively. For the sake of economy of presentation in an already overlong treatment, I present here only the results of the grade 6 to high school graduation six years later (this also allows us later to compare these results with data reported by TEA on grade 7-12 dropout rates). These are presented for cohorts labeled by their expected year of high school graduation. The cohort class of 1999, for example, would have been in grade 6 in 1992-93.

Figures 5.5 and 5.6 show the progress of grade 6 White and minority (Black and Hispanic) grade 6 cohorts of students to grades 8, 10, 11, 12 and high school graduation. As can be seen, over the last 20 years, for both White and minority cohorts, close to 100% of grade 6 students appear to be progressing to grade 8 two years later. For White students in grade 6 cohorts of the classes of 1982-85, about 90% proceeded to grade 11 and 12 on time and about 80% graduated six years after they were in grade 6. For minority grade 6 cohorts the rates of progress were lower: for grade 6 cohorts of the classes of 1982-85 about 80% of Black and Hispanic students progressed on time to grades 11 and 12 and about 65% graduated.

For classes of 1986 to 1990, there were slow but steady declines in all rates of progress for White students, from grade 6 to 8, from grade 6 to 10, etc. For minority cohorts of the classes of 1986 to 1990, there were initially sharper declines in rates of progress to grades 10, 11, and 12, but the cohorts of the 1989 and 1990 classes showed some rebounds in rates of progress to grades 10, 11 and 12 (and for the 1990 cohort to graduation). These patterns are associated with implementation of the first Texas high school graduation test, the TEAMS from 1985 to 1990.

In 1991, the initial year of TAAS testing, the grade 6 to high school graduation ratios fell precipitously; from 1990 to 1991, the ratio fell from 0.75 to 0.68 for Whites and from 0.65 to 0.55 for minorities. From 1992 to 1996, this ratio held relatively steady, for Whites at about 0.75 and for minorities at about 0.60. Since 1996, there have been slight increases in the high school graduation to grade six ratios, for Whites to 0.78 in 1999 and for minorities to almost 0.65.
Stepping back from specific numbers represented in Figures 5.5 and 5.6, three broad findings are apparent. First, the plight of Black and Hispanic students in Texas is not quite as bleak as it appeared when looking at grade 9 to high school graduation ratios, which showed only 50% since 1992 progressing from grade 9 to high school graduation. The bottom line in Figure 5.6 indicates that for most classes of the 1990s, 60-65% of Black and Hispanic students progressed from grade 6 to graduate on-time six years later (the grade 9 to graduation ratios are lower because of the increasing rates of retention in grade 9).

Figure 5.5 Progress from Grade 6 to Graduation, White Cohorts, Classes of 1982-2003

Figure 5.6 Progress from Grade 6 to Graduation, Minority Cohorts, Classes of 1982-2003
Second, one of the major features of both Figures 5.5 and 5.6 is that in each, the bottom two lines (representing the grade 12 to grade 6, and graduation to grade 6 ratios) tend to converge over the last 20 years. This means that over this period, given that students make it to grade 12, they are increasingly likely to graduate. For White students for example, in the class of 1999, almost 80% progressed from grade 6 to grade 12, and 78% to graduation. In contrast, in the classes of the early 1980s, around 90% were making it from grade 6 to grade 12, but only about 80% were graduating. For minority classes of the early 1980s, about 80% were progressing on-time to grade 12, but only about 65% graduating. For minority classes of 1998 and 1999, 68-69% progressed to grade 12 and 64-65% to graduation on time. In other words, a major pattern revealed in these two figures is that since high school graduation testing was introduced in Texas in the mid-1980s, one major change appears to have been that larger proportions of students who reach grade 12 do graduate.

The flip side of this pattern is that over this interval, smaller proportions of students, both White and minority are progressing as far as grade 12. For White classes of the early 1980s, about 90% of students in grade 6 progressed to grade 12 six years later, but by the 1990s the corresponding ratios had dropped to slightly below 80%. For minority classes of the early 1980s around 80% progressed from grade 6 to grade 12 six years later, but by the 1970s only 70% were progressing on time to grade 12.

The most obvious reasons for these substantial declines in progress from grade 6 to grade 12 six years later are increased rates of retention in grades before 12 and increased rates of dropping out before grade 12. In the next section, we review data on rates of retention in grade in Texas, and in Section 5.5 explain an alternative strategy to estimate numbers of dropouts.

5.4 Cumulative Retention Rates

In 1998, the TEA published the 1998 Comprehensive Biennial Report, containing statewide rates of retention in grade, reported by ethnicity. These data are of interest for several reasons. First, these data provide confirmation of what was apparent in the data shown in Figure 5.3, namely that the rate at which Black and Hispanic students are retained in grade 9 is 2.5 to 3.0 times that of the rate at which White students have to repeat grade 9.

### Table 5.1
Texas Statewide Rates of Retention in Grade 1996-97, by Ethnicity

<table>
<thead>
<tr>
<th>Grade</th>
<th>White % retained</th>
<th>Afric.-Amer % retained</th>
<th>Hispanic % retained</th>
<th>Total %</th>
</tr>
</thead>
<tbody>
<tr>
<td>K</td>
<td>2.30%</td>
<td>1.40%</td>
<td>1.60%</td>
<td>1.80%</td>
</tr>
<tr>
<td>1</td>
<td>4.40%</td>
<td>7.00%</td>
<td>6.60%</td>
<td>5.60%</td>
</tr>
<tr>
<td>2</td>
<td>1.60%</td>
<td>3.20%</td>
<td>3.40%</td>
<td>2.50%</td>
</tr>
<tr>
<td>3</td>
<td>0.90%</td>
<td>2.10%</td>
<td>2.10%</td>
<td>1.50%</td>
</tr>
<tr>
<td>4</td>
<td>0.70%</td>
<td>1.30%</td>
<td>1.40%</td>
<td>1.10%</td>
</tr>
<tr>
<td>Grade</td>
<td>White</td>
<td>Black</td>
<td>Hispanic</td>
<td></td>
</tr>
<tr>
<td>-------</td>
<td>-------</td>
<td>-------</td>
<td>----------</td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>0.60%</td>
<td>0.90%</td>
<td>1.00%</td>
<td>0.80%</td>
</tr>
<tr>
<td>6</td>
<td>1.00%</td>
<td>2.10%</td>
<td>2.30%</td>
<td>1.60%</td>
</tr>
<tr>
<td>7</td>
<td>1.60%</td>
<td>3.70%</td>
<td>3.80%</td>
<td>2.70%</td>
</tr>
<tr>
<td>8</td>
<td>1.30%</td>
<td>2.10%</td>
<td>2.90%</td>
<td>2.00%</td>
</tr>
<tr>
<td>9</td>
<td>9.60%</td>
<td>24.20%</td>
<td>25.90%</td>
<td>17.80%</td>
</tr>
<tr>
<td>10</td>
<td>4.80%</td>
<td>11.60%</td>
<td>11.40%</td>
<td>7.90%</td>
</tr>
<tr>
<td>11</td>
<td>3.20%</td>
<td>8.30%</td>
<td>7.90%</td>
<td>5.40%</td>
</tr>
<tr>
<td>12</td>
<td>2.50%</td>
<td>6.30%</td>
<td>7.20%</td>
<td>4.40%</td>
</tr>
<tr>
<td>Total</td>
<td>2.70%</td>
<td>5.70%</td>
<td>5.80%</td>
<td>4.20%</td>
</tr>
</tbody>
</table>


These data also allow us to see that despite much rhetoric lately about so-called "social promotion," retention in grade may be more common for Black and Hispanic students in Texas than is social promotion. Using an approach suggested by Robert Hauser, I analyzed data on patterns of retention in grade in Texas statewide as reported by the Texas Education Agency (and summarized in the table above). The approach suggested by Hauser is simply to subtract annual grade retention rates from 1.00 to yield rates of non-retention. The non-retention rates can then be multiplied across the grades to yield "compound" non-retention rates. The results for 1996-97 are shown in Table 5.2.

Table 5.2
Cumulative Rates of Grade Promotion, 1996-97

<table>
<thead>
<tr>
<th>Grade Level</th>
<th>White</th>
<th>Black</th>
<th>Hispanic</th>
</tr>
</thead>
<tbody>
<tr>
<td>Grades 1-3</td>
<td>93.22%</td>
<td>88.13%</td>
<td>88.33%</td>
</tr>
<tr>
<td>Grades 4-6</td>
<td>97.72%</td>
<td>95.76%</td>
<td>95.37%</td>
</tr>
<tr>
<td>Grades 7-8</td>
<td>97.12%</td>
<td>94.28%</td>
<td>93.41%</td>
</tr>
<tr>
<td>Grades 9-12</td>
<td>81.22%</td>
<td>57.57%</td>
<td>56.11%</td>
</tr>
<tr>
<td>All twelve grades</td>
<td>71.86%</td>
<td>45.81%</td>
<td>44.15%</td>
</tr>
</tbody>
</table>

Source: Based on TEA, 1998 Comprehensive Biennial Report, Table 4.2, p. 53.

White students have a probability of progressing through 12 grades without being retained in grade of about 72%. However, for Black and Hispanic students the comparable rates are 46% and 44%. In short, even before the end of so-called social promotion, Black and Hispanic students in Texas appear more likely than not to be retained in grade over the course of a 12-year school career. Note also that the compound retention rate for Hispanics (56%) is about double that for White students.
(28%), even before taking into account that Hispanics are much more likely than White students to drop out of school before grade 12. Note also that even before the secondary level of education, Black and Hispanic students in Texas are more likely not to be promoted (that is, to be retained in grade) than White students. The data in Table 5.2 indicated that at both the early elementary (grades 1-3) and upper elementary (grades 4-6) Black and Hispanic students are 70-75% more likely than White students to be "flunked," and retained to repeat a grade in school.

5.5 Dropouts and the Illusion of Progress

The retention rates shown in Table 5.1 may be used together with statewide enrollment data for 1995-96 and 1996-97 to calculate the grade levels at which students are dropping out of school in Texas. The logic of these calculations is as follows. If we assume no net migration of students into Texas, the number of students enrolled in say, grade 6 in 1996-97 ought to be equal to the sum of the number of students enrolled in grade 5 times the rate of non-retention in grade 5, plus the number enrolled in grade 6 times the grade 6 retention rate. Using this approach we may calculate the predicted grade enrollments in 1996-97 and compare them with the actual 1996-97 enrollments. Table 5.3 and Figure 5.7 show the results of such calculations for the Black, Hispanic, White and Total groups of students enrolled in Texas schools.

As can be seen, across all groups for grades 2 through 9 the enrollments for 1996-97 predicted on the basis of 1995-96 enrollments and reported rates of retention are quite close to the actual enrollments for 1996-97. For these grade levels the actual enrollments vary from those predicted by less than about 2%. For grade 1, actual enrollments in 1996-97 exceed those predicted by 5- 6%. The differences between actual and predicted grade 1 enrollments are fairly consistent across ethnic groups and presumably derive from the fact that across all groups kindergarten attendance was not universal in 1995-96 (hence the grade 1 enrollments in 1996-97 are larger than predicted from 1995-96 kindergarten enrollments).

Table 5.3
Grade Enrollments in Texas, 1996-97
Predicted and Actual Minus Predicted

<table>
<thead>
<tr>
<th>Grade</th>
<th>Black</th>
<th>Hispanic</th>
<th>White</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>1st</td>
<td>42,925</td>
<td>2,870</td>
<td>6.3%</td>
<td>117,564</td>
</tr>
<tr>
<td>2nd</td>
<td>42,998</td>
<td>917</td>
<td>2.1%</td>
<td>112,510</td>
</tr>
<tr>
<td>3rd</td>
<td>42,112</td>
<td>584</td>
<td>1.4%</td>
<td>109,434</td>
</tr>
<tr>
<td>4th</td>
<td>42,016</td>
<td>499</td>
<td>1.2%</td>
<td>107,748</td>
</tr>
<tr>
<td>5th</td>
<td>41,052</td>
<td>388</td>
<td>0.9%</td>
<td>105,989</td>
</tr>
<tr>
<td>6th</td>
<td>41,390</td>
<td>369</td>
<td>0.9%</td>
<td>106,360</td>
</tr>
<tr>
<td>7th</td>
<td>41,220</td>
<td>513</td>
<td>1.2%</td>
<td>105,656</td>
</tr>
<tr>
<td>8th</td>
<td>40,208</td>
<td>19</td>
<td>0.1%</td>
<td>104,465</td>
</tr>
<tr>
<td>9th</td>
<td>50,696</td>
<td>392</td>
<td>0.8%</td>
<td>131,492</td>
</tr>
<tr>
<td>10th</td>
<td>42,418</td>
<td>-5,791</td>
<td>-15.8%</td>
<td>103,814</td>
</tr>
<tr>
<td>11th</td>
<td>34,138</td>
<td>-3,504</td>
<td>-11.4%</td>
<td>79,894</td>
</tr>
<tr>
<td>12th</td>
<td>27,732</td>
<td>-1,679</td>
<td>-6.4%</td>
<td>64,911</td>
</tr>
</tbody>
</table>

Note however that for grades 10, 11 and 12 much larger disparities are apparent and vary considerably by ethnic group. Overall, enrollments in grades 10, 11 and 12 in 1996-97 were more than 65,000 lower than predicted based on the previous year's enrollments. The missing students were predominantly Black and Hispanic. Grade 10 enrollments in 1996-97 were about 16% lower than expected for Black and Hispanic students, but only about 8% lower than expected for White students.

What happened to these missing students? It seems extremely likely that they dropped out of school. This is not terribly surprising since previous research shows clearly that retention in grade is a common precursor to dropping out of school.

The grade 9 retention rates in Texas are far in excess of national trends. A recent national study, for example, showed that among young adults aged 16-24, only 2.4 percent had been retained in grades 9-12 (NCES, Dropout rates in the United States 1995, Report No. dp95/97473-5). The recent report of the National Research Council (NRC) also shows Texas to have among the highest grade 9 retention rates for 1992 to 1996 among the states for which such data are available (Heubert & Hauser, 1999, Table 6-1). (Note 14)

A casual observer might well wonder what is wrong with retaining students in grade 9 if they are academically weak. The answer is explained in the recent report on high stakes testing from the National Research Council:

The negative consequences, as grade retention is currently practiced, are that retained students persist in low achievement levels and are more likely to drop out of school. Low performing students who have been retained in kindergarten or primary grades lose ground both academically and socially relative to similar students who have been promoted (Holmes, 1989; Shepard
and Smith, 1989). In secondary school, grade retention leads to reduced achievement and much higher rates of school dropout. (Heubert & Hauser, 1999, p. 285).

Even the TEA has acknowledged that "research has consistently shown that being overage for grade is one of the primary predictors of dropping out of school in later years. . . . Being overage for grade is a better predictor of dropping out than underachievement." (TEA, 1996 Comprehensive Biennial Report on Texas Public Schools, pp. 35, 36.).

Hence, it is fair to say that the soaring grade 10 TAAS pass rates are not just an illusion, but something of a fraud from an educational point of view. Table 5.4 presents data to support this view.

### Table 5.4
**Texas Grade 10 Enrollments 1996-97 and Taking TAAS February 1997, by Ethnicity**

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Predicted</td>
<td>Actual</td>
<td>No.</td>
<td>% passing all 3 tests</td>
</tr>
<tr>
<td>Black</td>
<td>42,418</td>
<td>36,627</td>
<td>27,451</td>
<td>48.0%</td>
</tr>
<tr>
<td>Hispanic</td>
<td>103,814</td>
<td>88,845</td>
<td>69,421</td>
<td>52.0%</td>
</tr>
<tr>
<td>White</td>
<td>141,855</td>
<td>131,150</td>
<td>108,215</td>
<td>81.0%</td>
</tr>
</tbody>
</table>


What these data show is that the dramatically improved pass rates on the 1997 grade 10 TAAS tests are in part a result of students who dropped out (or are otherwise missing) between grade 9 in 1996 and the TAAS testing in February 1997. The overall pass rates reported by TEA on the 1997 grade 10 TAAS tests, of 48%, 52% and 81% for Black, Hispanic and White students, respectively, drop to 36%, 40.6% and 66.8% if we base the pass rates on the Fall 1996 actual enrollments. And they drop even further, to 31.1%, 34.8% and 61.8% if we base the pass rates on the number of students predicted to have been in grade 10 in 1996-97 (based, as explained above on the 1995-96 grade 9 enrollments and the TEA reported rates of retention in 1996-97).

This is, of course, also a reminder of an elementary fact of arithmetic. One can increase a proportion (such as percent passing) not just by increasing the numerator—but also by decreasing the denominator. In the next two sections, I estimate the proportions of the apparent gains in pass rates on the grade 10 TAAS tests between 1994 and 1998 that are attributable to decreases in the denominator (because of exclusion of students because either they dropped out of school or were classified as special education) and increases in the numerator (that is actual increases in numbers of students passing TAAS). Later, in Part 7, I return to the topic of dropouts in Texas, specifically to review and try to
reconcile sources of evidence about high school completion in Texas.

5.6 Increase in Special Education Exclusions

Before trying to distinguish the proportions of apparent TAAS gains that are real from those that are illusory, it is necessary to explain another manner in which students may be excluded from the grade 10 TAAS results used to rate secondary schools in Texas. It may be recalled that the soaring pass rates on the grade 10 TAAS summarized in Part 3 above were based on grade 10 students "not in special education." As far as I know, the TEA has not reported directly numbers of grade 10 students over time who were "in special education." However, TEA has reported the grade 10 pass rates separately for all students and for all students not in special education (at www.tea.state.tx.us/student.assessment/results/summary/). This allows us simply to subtract the two sets of data to derive the numbers and percentages of students who took the grade 10 TAAS who were classified as "in special education." Summary results are shown in Table 5.5. (Note 15)

| Table 5.5 | Number and % of Grade 10 TAAS Takers in Special Education, 1994-1998 |
| --- | --- | --- | --- | --- |
| **Numbers of Grade 10 TAAS Takers in Special Education** | Year | All groups | Afric.-Amer. | Hispanic | White |
| 1994 | 7602 | 833 | 1991 | 4685 |
| 1995 | 9049 | 1032 | 2351 | 5581 |
| 1996 | 11467 | 1500 | 3017 | 6810 |
| 1997 | 13005 | 1518 | 3707 | 7617 |
| 1998 | 14558 | 1818 | 4271 | 8284 |
| **Percentages of Grade 10 TAAS Takers in Special Education** | Year | All groups | Afric.-Amer. | Hispanic | White |
| 1994 | 3.9% | 3.3% | 3.3% | 4.5% |
| 1995 | 4.5% | 4.0% | 3.7% | 5.2% |
| 1996 | 5.3% | 5.4% | 4.5% | 6.1% |
| 1997 | 5.8% | 5.3% | 5.1% | 6.6% |
| 1998 | 6.3% | 6.3% | 5.7% | 7.1% |

As can be seen, the numbers and percentages of students taking the grade 10 TAAS, but classified as "in special education," have increased steadily between 1994 and 1998. This means that increasing numbers of students who have made it to grade 10 and taken the grade 10 TAAS have been excluded from school accountability ratings. Indeed between 1994 and 1998, the numbers of Black and Hispanic students taking the grade 10 TAAS counted as "in special education" more than doubled, though the percentage of
White students counted as "in special education" remained higher (7.1% vs. 6.3% and 5.7% for Black and Hispanic tenth graders, respectively). This means that a portion of the increase in pass rates on the grade 10 TAAS is attributable simply to the increases in the rates at which students were counted as in special education and hence excluded from school accountability ratings and from summary statistics showing pass rates for students not in special education.

5.7 How Much Illusion from Exclusion?

In Part 2 above, I reviewed evidence of the dramatic gains made in the passing rates on grade 10 TAAS between 1994 and 1998. As shown in Table 3.1, the percentage of students in Texas not in special education who passed all three grade 10 TAAS tests increased from 52% in 1994 to 72% in 1998, a 20 point increase. In the preceding two sections (5.5 and 5.6), we have seen that portions of this gain are purely an illusion due to increases in the numbers of students dropping out of school before taking the grade 10 TAAS, or else taking the grade 10 TAAS but excluded from accountability results because they are counted as "in special education." Hence, it is useful now to try to estimate what portion of the increased pass rate on TAAS is purely an illusion produced by these two kinds of exclusion.

In the previous section we saw that the percentage of students taking the grade 10 TAAS who were classified as "in special education" increased from 3.9% in 1994 to 6.3% in 1998. This suggests that around 2% of the 20-point gain in TAAS scores over this interval may be attributable simply to the increase in special education classifications. Note also that the increase in special education classifications has been larger for Black than for White students, so this may also account for a portion of the closing of the "race gap" in TAAS scores over this period. In contrast, the increase in Hispanic students classified as special education has been slightly less than the comparable increase for White students, so this factor could not account for the apparent shrinkage in the race gap in TAAS scores between Hispanic and White students.

What about the possible effects of increases in dropout rates in inflating the apparent grade 10 pass rates? To answer this question we would need to have estimates of the dropout rates between the early 1990s and 1998. In Section 5.5 above, I presented estimates of dropouts for one year, namely 1996-97. Nonetheless, the grade 8 to 9 progression ratios discussed in Section 5.2 clearly suggest that dropout rates increased between the early and late 1990s. Specifically between the early and late 1990s, the grade 8 to 9 progression ratios for Black and Hispanic students increased from around 1.20 to nearly 1.30. This suggests that the rate at which Black and Hispanic students are being retained in grade, and having to repeat grade nine increased over this interval by around 50%. Since grade retention is a common precursor to dropping out, this certainly suggests an increased dropout rate. At the same time, the analyses of progress for grade 6 cohorts presented in Section 5.3 revealed that grade 6 to grade 11 progression ratios for Whites and minorities varied by not more than 5% during the 1990s (for Whites, the ratio was consistently between 85% and 89%; and for minorities between 75% and 80%). The reason for focusing here on progress to grade 11 is because the data on enrollments is from the fall whereas TAAS is taken in the spring. But if students progress to grade 11, they presumably have taken the exit level version of TAAS in spring of the tenth grade. What this suggests is that the majority of the apparent 20-point gain in grade 10 TAAS pass rates cannot be attributed to exclusion of the types just reviewed. Specifically, if rates of progress from grades 6 to grade 11 have varied by no more than 5% for cohorts of the classes of the 1990s, this suggests even if we take this as an upper bound, the illusion attributable to dropping out is around 10% of the 20 point gain in TAAS scores.
bound, the extent to which increased retention and dropping out before fall grade 11, and add 2% for the increased rate of grade 10 special education classification, we still come up with less than half of the apparent 20-point gain in grade 10 TAAS pass rates between 1993 and 1998. So at this point in our analysis, it appears that while some of the gains may be due to these three forms of exclusion, a majority portion of the apparent gain is not. Hence it will be useful to turn in Part 7 to see whether the apparent gains on TAAS show up in any other evidence on the status and progress of education in Texas. Before turning to that topic, in Part 6 we review evidence from survey research about the effects of TAAS on education in Texas.
6. Educators' Views of TAAS

When it was learned in early May 1999 that the trial in the GI Forum case was to be postponed from June until September 1999, I realized that this delay would allow sufficient time to undertake surveys of Texas teachers about TAAS. We had a variety of indirect evidence that raised doubts about the validity and reliability of TAAS scores and the relationship of TAAS tests to secondary school teaching: TAAS results statewide from 1990 to 1998, the historical record concerning the setting of the passing score of 70% correct on the TAAS, patterns of grade enrollments in Texas over the last two decades, and data on the relationship between high school course grades and TAAS scores. However, we did not have any systematic evidence from those most directly affected by the TAAS graduation test, namely, secondary teachers and students, as to the educational value and effects of the TAAS testing. Consequently, a survey was in order.

Though we lacked the time and resources to survey the opinions of Texas students, with the help of Boston College graduate students, I undertook two different surveys of statewide samples of secondary teachers in Texas. One survey, previously described above was the survey for the Risk Analysis study. The second and larger survey was a "Survey of Testing and Teaching in Grades 7-12 in Texas," or what in shorthand was called the Testing and Teaching (TT) survey.

To undertake these surveys, I purchased mailing labels for a random sample of 4000 secondary teachers in Texas (specifically math and English/Language Arts) from Market Data Retrieval of Shelton, Connecticut. The number of 4,000 mailing labels was selected simply to meet the minimum purchasing requirements of this firm. From this list of 4,000, I then randomly selected 1,000 names to be used for the Testing and Teaching survey and 500 names to be used for the Risk Analysis survey. The survey forms were mailed on May 23, 1999, with self-addressed, stamped return envelopes. We tabulated all responses that were returned by the end of June, 1999, specifically 148 responses for the TT survey and 66 for the Risk Analysis survey. For both surveys we double-checked the accuracy of data entry before tabulating results. Since the Risk Analysis survey has been described in Part 4 above, I do not discuss it further here.

After undertaking this survey, I learned of two other surveys of Texas educators regarding TAAS: one by James Hoffman of the University of Texas at Austin and colleagues and the other by Gordon and Reese (1997). I describe these surveys in the order in which they were undertaken and reported, rather than the order in which I learned of them.

6.1 Survey/Interviews with Public School Teachers in Texas

Gordon and Reese surveyed 100 Texas teachers and followed up with interviews with 20 of the initial respondents. The authors do not explain how the survey
respondents were sampled, but they do mention they were "graduate students in educational administration" (Gordon & Reese, 1997, p. 349). Given the authors' affiliation as professors of educational administration at Southwest Texas State University, one suspects that respondents may well have been an opportunity sample of graduate students in the authors' program and perhaps other similar graduate programs. Nonetheless the authors do report that respondents' schools represented a cross section of Texas public schools relative to education level (elementary, middle school, high school), size, location (urban, suburban, rural), socioeconomic status (high and low SES) and TAAS category (exemplary, recognized, acceptable and low performing) (Gordon & Reese, 1997, p. 349).

In both the written survey and the follow-up interviews respondents were asked to address four broad questions:

1. How are students at your school prepared for TAAS?
2. What are the effects of TAAS on your students?
3. What are the effects of TAAS on you as a teacher?
4. What are the effects of TAAS on your school?

In the initial written survey, respondents were given a full blank page to respond to each question. In the follow-up interviews, a stratified random sample of respondents was chosen for in-depth interviews. Interviews were audio-taped, transcribed and coded to identify patterns among responses.

Regarding preparation for TAAS, respondents indicated that a huge amount of school time was devoted to coaching students for TAAS, with TAAS preparation becoming "all-consuming" during a period of four to eight weeks before the testing (p. 355). In most schools TAAS practice quizzes were administered on a regular basis with emphasis on teaching to the TAAS format, such as having students practice "bubbling" in answers on machine scorable answer sheets.

Respondents' answers regarding effects of TAAS on students were categorized as dealing with emotional, academic and social effects. For one group of students, teachers reported "no emotional effects at all because these students fail to recognize the importance of TAAS and are totally indifferent about it" (p. 356). A second group experiences moderate stress "which tends to motivate them to work harder to prepare for the test" (p. 356). A third group of students experience high stress as a result of TAAS. Among some in this group, according to respondents, the stress leads to anxiety and even panic. Among others it leads to anger and resentment. And another "subgroup eventually responds to the stress by "shutting down"; they cope by telling themselves they have no chance of doing well on TAAS and giving up" (p. 356). One respondent reported that the stress of TAAS "contributes to the dropout rate" (p. 357).

Regarding effects of TAAS on teachers, the vast majority of interview respondents (17 of 20) reported that TAAS leads to an emphasis on teaching TAAS-related content and "de-emphasis on teaching content not related to TAAS," (p. 359), including less emphasis on higher-level skills. All 20 interviewees also reported "concern, frustration and disappointment, caused by observing the negative effects of TAAS failure on at-risk students" (p. 360). Interviewees also reported that TAAS scores are "not accurate measures of the academic progress that their at-risk students have made" (p.360).

Regarding effects on teachers, "Nineteen of the interviewees agreed that TAAS makes them accountable in terms of teaching TAAS-related content, but that it does not make them accountable in terms of being effective teachers" (p. 360). While acknowledging the need for teacher accountability, respondents felt that TAAS was not a
good vehicle for achieving accountability because "TAAS is not a true measure of student learning and...it is unfair to use a single instrument like TAAS to compare the performance of teachers who are working with students of widely varying socioeconomic backgrounds, academic abilities and motivational levels" (pp. 360-61).

Regarding effects on schools, interviewees reported that "considerable human and material resources are expended on TAAS preparation" (p. 361) and that aspects of the curriculum that did not relate to TAAS were de-emphasized. Respondents were split as to whether or not their schools were "receiving pressure from parents and the community to do well on TAAS" (p. 362).

In their discussion, Gordon and Reese write that teacher respondents "reported not just 'teaching to the test' but also teaching to the test format, and doing so at the expense of large portions of the curriculum" (p. 363, emphasis in original). They also report that via focused "TAAS prep" teachers can "teach students how to respond correctly to test items even though the students have never learned the concepts on which they are being tested" (p. 364). The authors conclude that "drill and kill" coaching and preparation for TAAS are taking a "toll on teachers and students alike" and comment:

The most devastating effects of high-stakes testing seem to be occurring to the students who these tests are supposed to help the most—lower achieving students. Presumably, by setting clear standards and measuring results, state mandated tests make schools accountable for the basic education to which all children are entitled. According to participants in our study, however, their at-risk students' academic progress is being hindered by the negative effects of failing a test that many teachers insist does not measure what their students need to learn at their current stages of development, does not measure the progress their students have made, and is culturally biased. (Gordon & Reese, 1997, pp. 364-65).

The authors concluded with a number of recommendations for public dialog about the merits of high stakes testing, staff development, monitoring of the effects of high stakes testing, and establishment of a broader system of student assessment.

### 6.2 Testing and Teaching Survey of Secondary Math/English Teachers

I did not learn of the Gordon & Reese (1997) survey until recently. However, as previously explained, when the TAAS trial was postponed from June until September 1999, my colleagues and I decided to undertake a survey of a representative sample of teachers in Texas statewide.

The purpose of our Testing and Teaching survey was to obtain the opinions of a representative sample of secondary math and English/Language arts teachers in Texas statewide about the relationships between mandated testing and teaching and the effects of mandated testing. The survey form we used is a minor revision of a survey instrument that was administered to teachers nationwide in the early 1990s as part of a study funded by the National Science Foundation (Madaus et al., 1992). Specifically, from their survey instrument, one set of questions related to elementary education was deleted, one question was added, and space was provided at the end of the survey form for respondents to comment and provide their name and address, if they wished to receive a summary of survey results. Note that our survey form did not specifically ask about TAAS. A copy of our Testing and Teaching survey form is provided in Appendix 1.
By the end of June 1999, we had received 148 responses to our Testing and Teaching survey (representing a 14.8% response rate). (Note 16) After survey forms were received, data were entered and checked for accuracy, and a code book documenting data coding was developed. Before summarizing overall results of the Testing and Teaching survey, I should mention that on two of the forms returned, respondents had not completed answers to most questions, so they were excluded, leaving the main analysis sample with 146 respondents.

Respondents showed a good distribution of grade levels from 7 to 12, with several indicating teaching at more than one grade level. The vast majority were certified teachers (143) and roughly half (72) indicated that they had more than 12 years of teaching experience. The vast majority (123) also reported that they were "very comfortable" teaching their subject area.

As the survey form we used was addressed to the topic of mandated testing, it did not ask respondents directly about TAAS. However, in response to one question (C1), 118 respondents indicated that students in their class were required by their state or district to take standardized tests in the subject during the current calendar year. Space was provided for respondents to write the names of mandated tests to which they were referring and 112 respondents explicitly mentioned TAAS.

In response to a question about how mandated test results are used, respondents indicated that the most common uses were:

- to publish test scores (81%);
- to evaluate teachers (66%);
- to place students in programs (57%);
- to promote/graduate students (53%).

In contrast, only a minority of respondents (46%) indicated that mandated test results were used to alter the school curriculum.

In response to two sets of questions about teachers' own use and administrators' use of mandated test results, teachers indicated that results were "minimally" to "somewhat" important for a variety of purposes; but the uses rated most important across both sets of questions were two uses by administrators, namely school evaluation and district evaluation (both rated on average between "very" and "extremely" important).

A section of questions asked about test preparation. Results for these questions suggested a huge amount of test preparation, with the majority of respondents indicating that they do many different kinds of test preparation and 50% of respondents indicating that they spend more than 30 hours per year on test preparation. Also, 75% of respondents said that they begin test preparation more than one month before the mandated test.

In a set of questions addressed to the relationships between testing, curriculum and evaluation, respondents indicated that mandated testing influences teaching in a variety of ways, including influencing the increase or decrease of emphasis on certain topics and the content and format of tests that teachers use. In response to a question about the similarity of content of mandated testing and their own instruction, only 52% of 129 respondents answered "quite" or "very" similar.

Another series of questions asked about more general influences of mandated testing. The percentages of teachers agreeing (that is, agreeing or agreeing strongly) with each of these statements are summarized in Table 6.1.
Summary Results of Secondary Teachers' Answers about General Influences of Mandated Testing in Texas

<table>
<thead>
<tr>
<th>Statement</th>
<th>Percent of teachers answering &quot;Agree&quot; or &quot;Strongly Agree&quot; (n=139 to 142)</th>
</tr>
</thead>
<tbody>
<tr>
<td>6. Mandated testing influences teachers to spend more instructional time in whole group instruction.</td>
<td>65%</td>
</tr>
<tr>
<td>7. Mandated testing influences teachers to spend more instructional time in developing critical thinking skills.</td>
<td>45</td>
</tr>
<tr>
<td>8. Mandated testing influences teachers to spend more instructional time on individual seat work.</td>
<td>57</td>
</tr>
<tr>
<td>9. Mandated testing influences teachers to spend more instructional time in developing basic skills.</td>
<td>73</td>
</tr>
<tr>
<td>10. Mandated testing influences teachers to spend more instructional time with small groups of students working together (cooperative learning).</td>
<td>24</td>
</tr>
<tr>
<td>11. Mandated testing influences teachers to spend more instructional time solving problems that are likely to appear on tests.</td>
<td>88</td>
</tr>
<tr>
<td>12. Mandated testing influences teachers to spend more instructional time in the use of manipulatives and/or experiments for concept development.</td>
<td>22</td>
</tr>
<tr>
<td>13. Teachers in my district are gearing their instruction to mandated tests.</td>
<td>82</td>
</tr>
<tr>
<td>14. Mandated testing helps students achieve the objectives of the curriculum.</td>
<td>32</td>
</tr>
<tr>
<td>15. Teachers in my district have a pretty good idea of what students can do without using mandated tests.</td>
<td>82</td>
</tr>
<tr>
<td>16. The evaluation of teachers' competence is influenced (directly and/or indirectly) by their students' mandated test scores.</td>
<td>68</td>
</tr>
<tr>
<td>17. Mandated testing contributes to the realization of the goals of the current educational reform movement.</td>
<td>29</td>
</tr>
<tr>
<td>18. My state or district testing program sometimes leads teachers to teach in ways that go against their own ideals of good educational practice.</td>
<td>64</td>
</tr>
<tr>
<td>19. My district is putting pressure on teachers to improve their students' mandated test scores.</td>
<td>86</td>
</tr>
<tr>
<td>20. Students' mandated test scores are below the expectations of my school or district.</td>
<td>38</td>
</tr>
<tr>
<td>21. Mandated testing influences some teachers in my district to engage in non-standard testing practices (such as changing responses or increasing testing time limits).</td>
<td>12</td>
</tr>
</tbody>
</table>
While far more could be said about these results, key findings are as follows:

- Teachers in Texas are clearly feeling pressure to raise TAAS scores (86% of respondents agreed with the statement "My district is putting pressure on teachers to improve their students' mandated test scores.")
- Teachers have a pretty good idea of what students can do without mandated tests (82% agreed with the statement "Teachers in my district have a pretty good idea of what students can do without using mandated tests.")
- More teachers disagreed (45%) than agreed (32%) with the statement that "Mandated testing helps students achieve the objectives of the curriculum."
- More teachers disagreed (39%) than agreed (29%) with the statement that "Mandated testing contributes to the realization of the goals of the current educational reform movement."
- On the brighter side, results of the Testing and Teaching survey suggest that teachers and administrators are not widely engaging in non-standard testing practices (only 12% of respondents agreed with the last two statements (#21 and #22) in part F of the survey form). Indeed, one respondent commented "Perhaps I misunderstood questions 21 & 22. Are you asking if my district condones cheating? Absolutely not, the repercussions for that are very severe in this state" [173].

As indicated, the last portion of the Testing and Teaching survey form provided space for respondents to offer comments after these instructions: "If you would like to offer any comments about the relationship between mandated testing and teaching in Texas secondary schools, please write them here." A total of 51 respondents offered comments. On balance, these spontaneous comments on the relationship between mandated testing and teaching in Texas secondary schools were far more negative than positive about the role of mandated testing, with comments such as the following:

TAAS results haven't had the desired effect. It is used more as a "HAMMER" rather than a tool to improve. (Case 17)

I am not against mandated testing; but every time we work out a procedure for balancing the teaching, the state moves the test to a different grade level. We have it working well now, and now they're talking about moving it to 9th & 11th instead of 10th. (Case 39)

Mandated state TAAS Testing is driving out the best teachers who refuse to resort to teaching to a low-level test! (Case 67)

In citing these few comments here, I note that the full set of all respondents' comments appears as Appendix 2.

**6.3 Survey of Texas Reading Specialists**

The third survey of educators in Texas about TAAS was by Hoffman, Pennington & Assaf of the University of Texas—Austin and Paris of the University of Michigan.
did not learn of this survey until just before the TAAS trial in the Fall of 1999 and results of this survey were not allowed to be entered as evidence in the TAAS case. Nonetheless, Hoffman and colleagues have been very generous in sharing with me not just a manuscript reporting on their survey results, but also an entire set of their original data.

The Hoffman et al. (1999) survey was of members of the Texas State Reading Association (TSRA), an affiliate of the International Reading Association, whose membership includes classroom teachers, reading specialists, curriculum supervisors, and others in leadership positions. The purpose of the survey "was to examine the ways in which TAAS affects teachers, teaching and students from the perspective of the professional educators who are closest to classrooms and schools" (Hoffman et al., 1999, p. 3). The survey form contained 113 items, many duplicated or slightly adapted from Urdan & Paris's (1994) survey of teachers in the state of Michigan and the Haladyna, Nolen, and Haas (1991) survey of teachers in Arizona. The survey items were mostly Likert-scale items (with a five-point scale answer format: 1=strongly disagree, 2=disagree, 3= agree, 4=strongly agree, and 5=don't know) asking about attitudes, test preparation and administration practices, uses of scores, effects on students, and overall impressions of trends. In addition, five items were included containing invitations for extended written responses.

The authors surveyed a random sample of members of the TSRA. After a reminder letter and a second random sampling, they received a total of 201 usable responses representing an overall return rate of 27% of surveys sent (representing 5% of the total membership of TSRA). The authors report that no biases were detected in the response rates "based on geographical areas of the state" of Texas (p. 4).

The authors reported results in three different ways: percentages responding to particular questions in particular ways, scaled response representing answers summed across items relating to similar topics, and verbatim quotations of written responses. Overall, respondents to the Hoffman et al. survey were older (61% between the ages of 40-60), and more experienced (63% with over 10 years experience and 45% with over 20 years experience) than classroom teachers in general in the state of Texas (p. 5).

Scaled score responses indicated that on a composite measure of general attitudes toward TAAS "reading specialists strongly disagree with some of the underlying assumptions and intentions for TAAS" (p. 5). Other scaled score responses revealed that that "reading specialists challenge the basic validity of the test and in particular for minority and ESL speakers who are the majority in Texas public schools" (p. 6). Another composite variable representing general attitudes towards TAAS reflected "a strong negative attitude toward TAAS" (p. 7).

Respondents' answers regarding effects of TAAS on students revealed that a majority said that TAAS often or always caused student irritability, upset stomachs and headaches. Responses to three questions regarding overall impressions of TAAS were particularly striking. One question asked:

The results from TAAS testing over the past several years seem to indicate that scores are on the rise. Do you think this rise in test scores reflects increased learning and higher quality teaching?

To this question, 50% answered no, and 27% answered yes.

Another question read as follows:

It has been suggested that the areas not tested directly on TAAS (e.g., fine
arts) and other areas not tested at certain grades levels (e.g., science at the 4 th grade level) receive less and less attention in the curriculum. What do you feel about this assertion?

In response to this question, 85% answered "very true" or "somewhat true." A third question read as follows:

It has also been suggested that the emphasis on TAAS is forcing some of the best teachers to leave teaching because of the restraints the tests place on decision making and the pressures placed on them and their students.

A total of 85% of respondents agreed with this statement.

Written comments "revealed the depth of feeling and passion on the part of teachers with respect to trends in TAAS testing:"

I am very sad that education has stooped to the low level of measuring performance with standardized testing and Texas has taken it even lower with their TAAS. We know what works in education. We just seem to ignore the research and keep on banging our heads against the "TAAS wall" and "retention walls."

Please support teachers more than ever. Our children are hurting more than ever. If there was ever a time to change it is now. Give teachers back their classrooms. Let them teach and spend quality time with their students. They need us!

I think TAAS is the biggest joke in Texas. I have never seen such an injustice.

I believe that TAAS interferes with the very nature of our job. The pressure from administrators to increase campus scores leaves teachers little time for real instruction...."

My heart breaks to see so many teachers "just surviving." I believe that our solution is just to support each other because the public has no real concept of the situation.

TAAS is ruining education in Texas! Help!

6.4 Similarities and Differences in Survey Results

The surveys summarized above were undertaken independently and polled somewhat different samples of Texas educators. Gordon and Reese surveyed Texas teachers who were "graduate students in educational administration" (Gordon & Reese, 1997, p. 349). Though the authors do not explain exactly when this survey was conducted, it was presumably around 1996. The survey by Hoffman et al. and the one undertaken by me were both performed during 1998-99, though of somewhat different populations. Hoffman et al. surveyed reading specialists statewide, while I surveyed secondary math and English/language arts teachers. Despite these differences, results of the three independent surveys of Texas educators have four broad findings in common.

Texas schools are devoting a huge amount of time and energy preparing
**students specifically for TAAS.** As mentioned, in the Gordon & Reese survey, respondents reported a huge amount of school time was devoted to coaching students for TAAS, with TAAS preparation becoming "all-consuming" during a period of four to eight weeks before the testing (p. 355). In the Testing and Teaching survey, 75% of respondents said that they begin test preparation more than one month before the mandated test (TAAS). And in the Hoffman et al. survey, when asked whether the rise in TAAS scores reflected "increased learning and higher quality teaching," nearly twice as many respondents answered "no" (50%), as answered "yes" (27%). In their written comments to this question many teachers explained that they felt test preparation was what accounted for the rising scores:

I feel that it reflects that we are doing a better job teaching for the test. We are being forced to teach the test. (Case 11).

Students are being trained earlier on how to take the TAAS test. In 5-10 more years a different format will be provided & low scores will be the reason to teach to that test too. (Case 17).

I think students know how to take the test because we practice ad nauseum. (Case 20).

TAAS is a poor measure of actual student performance. Increases are due to becoming accustomed to the test. (Case 38).

Teachers are teaching to TAAS period. Curriculum is directed by TAAS even in K. TAAS doesn’t address all areas - if it did colleges would have better results than ever before instead of remedial classes! (Case 46).

The scores reflect an increase in time spent on one test instead of teaching students the regular curr.[iculum]. (Case 49).

Teachers are spending the school day teaching to the test. (Case 84).

Higher quality teaching is not exhibited on a daily basis. However TAAS test taking skills occur everyday. (Case 95).

We've been teaching to the TAAS so long, our students are used to it. (Case 102).

The rising scores may be a result of better test-taking skills rather than knowledge. (Case 131).

No, School districts have figured out how to teach to the TAAS and to exclude students from being accounted. (Case 136).

I believe the scores reflect that students are learning test-taking strategies. (Case147).

Kids are just being programmed on how to take and pass the TAAS test, not truly mastering skills. (Case 151).

Teachers are learning how to teach the TAAS (Case 169).
I believe that students are simply being taught to take the test, not learn and apply the knowledge. (Case 199).

One comment from the Hoffman et al. survey described the emphasis on test preparation this way:

Our campus has 2 practice TAAS (annually) (Nov. & Feb) plus the "real" taas. Our weekly lesson plans contain TAAS warm-ups, TAAS lesson objectives, and 20 min of reading. I personally am sick of TAAS by April & May. My Teacher evaluation last yr was down because my student scores were down by 7 pts. I personally have 6 friends who quit teaching altogether because of TAAS. (Case No. 94)

Even some of the teachers who answered "Yes" in the Hoffman survey, that the rise in TAAS scores did reflect "increased learning and higher quality teaching," qualified their answers considerably in their written comments:

Students are learning more of the basic skills TAAS tests because teachers are figuring out better ways to teach them. Students are NOT receiving a well-rounded education because Social Studies & Science are being cut to teach TAAS skills. (Case 106).

Yes, there is increased learning but at a partial price. I have seen more students who can pass the TAAS but cannot apply those skills to anything if it's not in TAAS format. I have students who can do the test but can't look up words in a dictionary and understand the different meanings. They can write a story but have trouble following directions for other types of learning. As for higher quality teaching, I'm not sure that I would call it that. Because of the pressure for passing scores, more and more time is spent practicing the test and putting everything in TAAS format. (Case 184).

A handful of respondents suggested that that the rise in TAAS scores was due not to test preparation or increased learning, but to the TAAS tests getting easier over time, to schools excluding low scoring students, or to administrators' cheating:

TAAS scores have seemed to rise in election years. The tests seemed easier in those years. (Case 121).

It seems as though the questions are actually easier. (Case 127).

No, School districts have figured out how to teach to the TAAS and to exclude students from being accounted. (Case 136)

I think the tests are easier to make the legislators look better. (Case 155).

The test seems to have gotten easier. (Case 159)

There are a lot of teachers and administrators who know how to "cheat" and get higher scores by kids. They don't want their school to score bad, so they cheat. (Case 160).

I also think there are admin. who are cheating ex. Austin schools. (Case
Emphasis on TAAS is hurting more than helping teaching and learning in Texas schools. As mentioned, the results of the Hoffman et al. survey showed that a clear plurality of respondents (50%) reported that TAAS score gains were not due to "increased learning and higher quality teaching." No directly analogous questions were asked in the Testing and Teaching or Gordon & Reese surveys, but some of the findings from these surveys confirm Texas teachers’ generally negative views about the educational impact of TAAS. Recall that in the Testing and Teaching survey, it was found that more teachers disagreed (45%) than agreed (32%) with the statement that "Mandated testing helps students achieve the objectives of the curriculum." Also, more teachers disagreed (39%) than agreed (29%) with the statement that "Mandated testing contributes to the realization of the goals of the current educational reform movement." Recall also that Gordon & Reese concluded that "drill and kill" coaching and preparation for TAAS were taking a "toll on teachers and students alike"—especially "lower achieving students" whose "academic progress is being hindered by the negative effects of failing a test that many teachers insist does not measure what their students need to learn at their current stages of development, does not measure the progress their students have made, and is culturally biased." (pp. 364-65).

As in the Hoffman et al. survey, written responses to our Testing and Teaching survey help to convey something of teachers' depth of feeling and passion about TAAS:

Texas has the "Texas Assessment of Academic Skills" test. Most schools have established a class strictly for the TAAS test. Our curriculum is based on previous TAAS test questions. We "teach the TAAS" in our classes. Our administrators have even gone as far as incorporating TAAS objectives and materials into daily instruction in ALL subject areas. We are not covering skills for higher level thinking at times because of state mandated tests. (Case 13).

Testing is now more important than teaching. Students learn much about testing, little about subject. (Case 20).

We are testing our students to death! My students have been taken out of class four times this year for standardized testing. Too MUCH! (and for what?) (Case 29).

There are too many loopholes. Students who were never on an IEP or in CM are being forced into it so that they will be exempt from standardized tests. (Case 42).

Mandated state TAAS Testing is driving out the best teachers who refuse to resort to teaching to a low level test! (Case 67).

TAAS has become the Be All and End All. It is ridiculous to put so much on one test, where even good students have been known to guess and not even read the question. I have seen them. Our school can be at risk because one student chooses to mess up. One year we were on probation for 1 student over the limit. (Case 87, emphasis in original).

Mandated testing has severely damaged the mathematics curriculum! (Case
Teaching to the TAAS results in a level of education which is substandard. I strongly feel TAAS should be abolished. (Case 104).

Mandatory tests are hard on both teachers and students. Our state set the End of Course test one week before semester finals. The stress level for all of us is high. The end of school in itself is difficult. Why do we compound the situation by adding another useless test. Our state is also taking the EOC out of the schedule. They are replacing the EOC with another TAAS test. At least the EOC covered current material. Now extra work is added because the TAAS covers different areas than Algebra essential elements. (Case 123).

We are so concerned about the TAAS & End of Course exam that we are teaching the test, but the kids are not learning the material. I can teach the test, and have a very high percentage pass, yet have kids that know no Algebra. Going to three years TAAS testing in the future will reduce education to completely teaching the test, and we will graduate an illiterate generation. (Case 130).

I really feel that we are definitely getting away from teaching the basic concepts to teaching the test and this is very sad because the farther the student goes in mathematics the less he or she knows of the why's. (Case 133).

It stifles professional growth and academic growth as well. Too much emphasis on testing. (Case 147).

**Emphasis on TAAS is particularly harmful to at-risk students.** A third finding common across the three surveys is that the focus on TAAS in Texas is especially harmful to particular kinds of students. This finding is interesting because none of the surveys asked directly about this issue. Nonetheless, the matter arose in all three inquiries. Recall Gordon & Reese's concluding comment that in the common opinion of their interviewees "their at-risk students' academic progress is being hindered by the negative effects of failing a test that many teachers insist does not measure what their students need to learn at their current stages of development, does not measure the progress their students have made, and is culturally biased" (Gordon & Reese, 1997, pp. 364-65). Also, spontaneous comments in the Testing and Teaching and Hoffman et al. surveys raised similar concerns. From the former:

I personally wonder about the fairness of these tests. Children from lower SES tend not to do as well. Therefore, it tends to be discriminatory I think. I think some children do not have the cultural experiences that help them answer the questions accurately. (Case 84).

The TAAS test is driving the curriculum and not teaching students how to think. It also punishes ESL students—they can complete four years of high school with adequate grades but not be allowed to walk at graduation because they do not have enough command of English to pass the TAAS Exit. (Case 66).
In teachers’ written comments in the Hoffman et al. survey, several teachers mentioned the problems created for special education students by emphasis on TAAS. Here is one extended example:

Special education assessments. . . and diagnostic evaluations are \textit{NOT} aligned with TAAS objectives. Therefore children are sometimes not qualified for spe. ed. services who have low IQs and yet are expected to pass TAAS to graduate. I.E.P. goals for reading and math (other than "mainstream" IEP's) are not compatible with TAAS in our district. Reading goals are not detailed enough in comprehension, math is not grade-level appropriate. IEP's tend to emphasize discrete skills, such as computation while TAAS emphasizes application and problem solving. Texas criteria for diagnosis of L.D. do not take into consideration TAAS standards. Teachers, under pressure to have good scores, over-refer students for spe. ed. testing—sometimes 1/3-1/2 of their classes! Most administrators (\textit{NOT} mine) pressure ARD committees to exempt all students in spe. ed. from taking TAAS. Appropriate alternative assessments are not available. TAAS does not take into account LEP students, or students in special education, who are being "included" in higher numbers. (Case No. 89) (Note 17)

\textbf{Emphasis on TAAS contributes to retention in grade and dropping out of school.}

Finally, all three surveys provide support for the proposition that emphasis on TAAS contributes to both retention in grade and students dropping out of school. One question in the Hoffman et al. survey asked respondents:

Are there efforts to exclude/exempt students from testing who might not do well on the test and thereby negatively affect a school's rating?

Overall, 67\% of respondents answered "often" or "sometimes" in response to this question. Obviously, there are ways of excluding students other than by retention in grade and encouraging drop outs (such as special education classification). But recall that one out of 20 interviewees in the Gordon and Reese survey said directly that "the stress of TAAS contributes to the dropout rate." A majority of respondents in the Testing and Teaching survey rated "to promote/graduate students" as a common use of mandated tests in Texas. Additionally, many written comments in the Hoffman et al. surveys expressed dissatisfaction with the practice of retaining students in grade based on TAAS scores, irrespective of other evidence about student learning.

In concluding this discussion of the results of three surveys of Texas educators regarding TAAS, it is only fair to add one major caveat. Despite the preponderance of negative comments about the effects of TAAS on education in Texas, there were some comments suggesting that the role of TAAS is at least somewhat beneficial:

It seems to work out fairly well for most of us with TAAS, however, the end of course tests are not that useful. (Case 5).

I believe there is a purpose for these tests. If nothing else, it gives teachers goals for their students. But I do not believe my teaching competence should be based on those scores \textit{solely}. (Case 34, emphasis in original).

In light of this contrast, with most teachers reporting the effects of TAAS to be harmful, but with a minority reporting positive effects, it is useful to draw back, to try to
gain a broader picture of the status of education in Texas. It is to such a perspective that we turn in Part 7. (Note 18)
7. Other Evidence on Education in Texas

Beyond the views of teachers, what other evidence is available that might provide a picture of the status and progress of education in Texas? In Part 7, we review four kinds of evidence. First, we compare sources of evidence on high school completion in Texas with the data previously presented in Part 5 above. Next we compare data on retention in grade for states which have reported such data. In Section 7.3 we review evidence available from SAT college admissions testing over the last 30 years. Then, in Section 7.4 we return to take a closer look at NAEP data—some of which, as we saw in Part 2 above, has previously been cited as evidence of the Texas "miracle" in education. Finally, we comment briefly on several other sources of evidence about education in Texas.

7.1 Dropout Data on Texas Revisited

As mentioned previously, when I first started studying education in Texas approximately two years ago, a major discrepancy quickly contributed to my suspicions about the validity of the TEA reported data on dropout rates in Texas (some of which was reproduced in Table 3.3 above). The TEA data showing declining dropout rates in Texas were contradicted by two independent sources of evidence: a series of attrition studies reported by the Intercultural Development Research Association (IDRA), and reports on dropouts in the United States from the National Center for Education Statistics (NCES). The IDRA and NCES sources did not, however, contain estimates of dropout rates for Texas as far back as I needed to examine the apparent effects of high school graduation testing on grade enrollments and high school graduation. Consequently, I sought to analyze data on Texas high school graduates and enrollments by grade going back to the mid-1970s. Nonetheless, having done so, it is now helpful to recount the IDRA and NCES reports' findings and to compare them with results previously presented. Before reviewing and comparing these sources, let me review TEA-reported dropout data in more detail than was done in Part 3 above.

**TEA Dropout Data.** In the Fall of 1999, the Texas Education Agency (TEA) released a report titled *1997-98 Report on Texas Public School Dropouts*. (The report was originally issued in September 1999, and in a revised edition in December.) The highlights of the report were as follows:
How many students drop out?

- In 1997-98, a total of 27,550 students in Grades 7-12 dropped out of Texas public schools.
- Statewide, the annual dropout rate was 1.6 percent, unchanged from 1996-97.
- The 1997-98 actual longitudinal dropout rate, calculated for a cohort of students tracked from 7th to 12th grade, was 14.7 percent.

Who drops out and why?

- About 77 percent of dropouts were overage for grade, down from over 80 percent in 1996-97.
- On average, males continued to drop out at a slightly higher rate than females.
- Hispanic students had the highest average dropout rate, at 2.3 percent, followed by African American students (2.1%).
- Reasons cited for dropping out of school included poor attendance, entering non-state-approved General Educational Development (GED) programs, and pursuing a job.

Are they leaving certain districts?

- School districts with the largest enrollments (50,000 or more students) had the highest average dropout rate, at 2.1 percent.
- Generally, districts with lower student passing rates on the Texas Assessment of Academic Skills (TAAS) had higher dropout rates.

How do we compare nationally?

- Based on the Current Population Survey, an estimated 4.6 percent of students in Grades 10-12 dropped out of school across the nation.
- Texas had one of the lower dropout rates out of 32 states that met required Common Core of Data collection standards for school year 1996-97. (TEA, 1999, 1997-98 Report on Texas Public School Dropouts, p. iii)

Table 8 of the TEA report presented data on "historical dropout rates by ethnicity." Figure 7.1 presents a graph of these data.
These data obviously indicate that the annual dropout rate in Texas (that is the numbers of dropouts reported in grades 7-12 divided by the grade 7-12 enrollment) has fallen dramatically in the last decade. I refrain from commenting further on these results until after summarizing other evidence on dropouts in Texas.

**IDRA Attrition Studies.** In the mid-1980s, under a contract with the Texas Department of Community Affairs, the Intercultural Development Research Association (IDRA), undertook a series of studies, one aim of which was to estimate "the magnitude of the dropout problem in the State of Texas" (IDRA, 1986, p. i). After describing the paucity of previous reliable research on dropouts in Texas, the IDRA researchers developed an index of attrition to estimate dropout rates not just statewide, but also at the level of school districts in Texas.

The index developed and used by IDRA consists of taking grade level enrollments for a base year and comparing them to enrollments in subsequent years. Since school and district enrollments are not constant, with changes in size due to increasing or declining enrollments, it is necessary to take the growth trend into account in computing attrition rate. The size change ratio was calculated by dividing the total district enrollment for the longitudinal study end year, by the total district enrollment for the base study year. Multiplying the base year enrollment by the district change ratio produces an estimate of the number of students expected to be enrolled at the end year. (IDRA, 1986, p. 9).

In short, the IDRA attrition index method for estimating dropouts is very similar to the way in which I calculated progress from grade 9 to high school graduation (as reported in Section 5.1 above). The IDRA method differs, however, in two respects from the one used in calculating results presented in Section 5.1. First, instead of simply assuming that the numbers of students in grade nine in a particular year (say 1990-91) in a particular school system represents a reasonable estimate of the numbers expected to graduate three years later (in 1993-94), the IDRA approach adjusts this estimate to take into account the overall growth or decline in enrollments in the system over the time period studied (thus, for example, if overall grade 9-12 enrollment increased 25%
between 1990-91 and 1993-94), the IDRA approach assumes that the number enrolled in grade 12 in 1993-94 would be 25% greater than the 1990-91 grade 9 enrollments). Second, the IDRA approach focuses on grade enrollments and has not been applied, at least insofar as I am aware, to the question of how many students actually graduate from Texas high schools at the end of grade 12.

The IDRA has regularly updated its attrition calculations since its original study in 1986. Table 7.1 presents the organization’s most recent results, showing percent attrition from grades 9 to 12, from 1985-86 to 1998-99 (note that data for 1990-91 are missing).

### Table 7.1

**IDRA Reported Attrition Rates, Grades 9-12 (% Attrition)**

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Black</td>
<td>34%</td>
<td>38%</td>
<td>39%</td>
<td>37%</td>
<td>38%</td>
<td>39%</td>
<td>43%</td>
<td>50%</td>
<td>51%</td>
<td>51%</td>
<td>49%</td>
<td>48%</td>
</tr>
<tr>
<td>White</td>
<td>27</td>
<td>26</td>
<td>24</td>
<td>20</td>
<td>19</td>
<td>22</td>
<td>25</td>
<td>30</td>
<td>31</td>
<td>32</td>
<td>31</td>
<td>31</td>
</tr>
<tr>
<td>Hispanic</td>
<td>45</td>
<td>46</td>
<td>49</td>
<td>48</td>
<td>48</td>
<td>49</td>
<td>51</td>
<td>53</td>
<td>54</td>
<td>53</td>
<td>53</td>
<td>53</td>
</tr>
<tr>
<td>Total</td>
<td>33</td>
<td>34</td>
<td>33</td>
<td>31</td>
<td>31</td>
<td>34</td>
<td>36</td>
<td>40</td>
<td>42</td>
<td>43</td>
<td>42</td>
<td>42</td>
</tr>
</tbody>
</table>


Comparison of the TEA and IDRA data reveals two broad findings. First, for the academic year 1988-89, their estimates of dropouts are somewhat comparable. For that year, the IDRA reported attrition rates of 37%, 20% and 48% for Black, White and Hispanic students respectively. And if we multiply the TEA-reported annual dropout rates for grades 7-12 by six to approximate a longitudinal dropout rate across this grade span, we get 45.1%, 27.3% and 48.6% for Black, White and Hispanic students respectively. These estimates are not terribly close, but at least they are in the same ballpark. And the differences are in the directions one would expect. The TEA reported data yield slightly higher percentages since they cover grades 7-12, while the IDRA attrition percentages cover just grades 9-12.

Second, after 1988-89, the IDRA and TEA results diverge dramatically. The IDRA data show attrition increasing between 1988-89 and 1998-99, while the TEA data show dropouts to be decreasing sharply over the same period. The divergence is so dramatic as to make one wonder whether the two organizations are referring to the same state—or even living on the same planet. The IDRA results show increases in attrition such that by 1997-98, 49% of Black, 31% of White and 53% of Hispanic students dropped out between grades 9 and 12. In contrast, the TEA reported data suggested longitudinal dropout rates for grade 7-12 of 12.6%, 5.4% and 13.8% for Black, White and Hispanic students respectively. In other words, the IDRA results indicate that the dropout problem in Texas in the late 1990s was four to six times worse than the TEA was reporting.

Whose estimates are to be trusted; those of the IDRA or of the TEA? Before giving my answer to this question, let me summarize results of one more organization, this one from outside Texas.

**NCES Dropout Studies.** Over the last decade the National Center for Education Statistics (NCES) has issued a series of reports on dropouts in the United States. The eleventh report in the series presents data on high school dropout and completion rates in
1998, and includes time series data on high school dropout and completion rates for the period 1972 through 1998. The high school completion rates are based on results of the Census Bureau's Current Population Surveys (CPS) of random U.S. households conducted in October of each year. The CPS surveys have not been designed with the specific intent of deriving state level high school completion rates and so in order to help derive reliable estimates, the NCES analysts who prepared the dropout reports have calculated averages across three years of CPS surveys. Also, it should be explained that the CPS data are based on self-reports of high school completion whether it be via normal high school completion or via alternative high school completion such as the GED testing. (Note 19)

Table 7.2 reproduces a table from the latest NCES dropout report, showing high school completion rates of 18 through 24 year olds, not currently enrolled in high school or below, by state: October 1990-92, 1993-95 and 1996-98. As can be seen for all three time periods, these data show Texas to have among the lowest rates of high school completion among the 50 states. In each time period, the median high school completion rate across the states was about 88%, while the completion rate for Texas was about 80%. This pattern indicates that the median noncompletion rate across the states is about 12% while that of Texas is about 20% (about 66% worse than the median of the other states).

**Table 7.2**

**High School Completion Rates of 18 Through 24 Year-olds, Not Currently Enrolled in High School or Below, by State: October 1990-92, 1993-95 and 1996-98**

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Total National</td>
<td>85.5%</td>
<td>85.8%</td>
<td>85.6%</td>
</tr>
<tr>
<td>Alabama</td>
<td>83.9</td>
<td>83.6</td>
<td>84.2</td>
</tr>
<tr>
<td>Alaska</td>
<td>86.9</td>
<td>90.5</td>
<td>88.3</td>
</tr>
<tr>
<td>Arizona</td>
<td>81.7</td>
<td>83.8</td>
<td>77.1</td>
</tr>
<tr>
<td>Arkansas</td>
<td>87.5</td>
<td>88.3</td>
<td>84.5</td>
</tr>
<tr>
<td>California</td>
<td>77.3</td>
<td>78.7</td>
<td>81.2</td>
</tr>
<tr>
<td>Colorado</td>
<td>88.1</td>
<td>88.4</td>
<td>85.5</td>
</tr>
<tr>
<td>Connecticut</td>
<td>89.9</td>
<td>94.7</td>
<td>91.6</td>
</tr>
<tr>
<td>D.C.</td>
<td>86.2</td>
<td>93.0</td>
<td>88.5</td>
</tr>
<tr>
<td>Delaware</td>
<td>84.0</td>
<td>87.7</td>
<td>84.9</td>
</tr>
<tr>
<td>Florida</td>
<td>84.1</td>
<td>80.6</td>
<td>83.6</td>
</tr>
<tr>
<td>Georgia</td>
<td>85.1</td>
<td>80.3</td>
<td>84.8</td>
</tr>
<tr>
<td>Hawaii</td>
<td>93.5</td>
<td>92.0</td>
<td>92.3</td>
</tr>
<tr>
<td>Idaho</td>
<td>84.7</td>
<td>86.1</td>
<td>85.8</td>
</tr>
<tr>
<td>Illinois</td>
<td>96.0</td>
<td>86.5</td>
<td>86.6</td>
</tr>
<tr>
<td>State</td>
<td>1st Year</td>
<td>2nd Year</td>
<td>3rd Year</td>
</tr>
<tr>
<td>-------------------</td>
<td>----------</td>
<td>----------</td>
<td>----------</td>
</tr>
<tr>
<td>Indiana</td>
<td>87.8</td>
<td>88.5</td>
<td>89.3</td>
</tr>
<tr>
<td>Iowa</td>
<td>94.6</td>
<td>93.2</td>
<td>88.0</td>
</tr>
<tr>
<td>Kansas</td>
<td>93.2</td>
<td>90.9</td>
<td>91.2</td>
</tr>
<tr>
<td>Kentucky</td>
<td>81.1</td>
<td>82.4</td>
<td>85.2</td>
</tr>
<tr>
<td>Louisiana</td>
<td>83.9</td>
<td>80.1</td>
<td>81.6</td>
</tr>
<tr>
<td>Maine</td>
<td>91.9</td>
<td>92.9</td>
<td>91.6</td>
</tr>
<tr>
<td>Maryland</td>
<td>88.6</td>
<td>93.6</td>
<td>94.5</td>
</tr>
<tr>
<td>Massachusetts</td>
<td>89.8</td>
<td>92.5</td>
<td>90.6</td>
</tr>
<tr>
<td>Michigan</td>
<td>87.2</td>
<td>88.6</td>
<td>91.0</td>
</tr>
<tr>
<td>Minnesota</td>
<td>92.5</td>
<td>93.1</td>
<td>90.0</td>
</tr>
<tr>
<td>Mississippi</td>
<td>85.4</td>
<td>93.9</td>
<td>82.0</td>
</tr>
<tr>
<td>Missouri</td>
<td>88.1</td>
<td>90.4</td>
<td>90.4</td>
</tr>
<tr>
<td>Montana</td>
<td>91.6</td>
<td>89.6</td>
<td>91.1</td>
</tr>
<tr>
<td>Nebraska</td>
<td>92.5</td>
<td>94.1</td>
<td>91.2</td>
</tr>
<tr>
<td>Nevada</td>
<td>82.1</td>
<td>81.9</td>
<td>78.2</td>
</tr>
<tr>
<td>New Hampshire</td>
<td>87.9</td>
<td>86.9</td>
<td>89.2</td>
</tr>
<tr>
<td>New Jersey</td>
<td>90.8</td>
<td>91.6</td>
<td>91.8</td>
</tr>
<tr>
<td>New Mexico</td>
<td>84.1</td>
<td>82.3</td>
<td>78.6</td>
</tr>
<tr>
<td>New York</td>
<td>88.0</td>
<td>87.0</td>
<td>84.7</td>
</tr>
<tr>
<td>No. Carolina</td>
<td>83.0</td>
<td>85.5</td>
<td>85.2</td>
</tr>
<tr>
<td>North Dakota</td>
<td>96.3</td>
<td>96.4</td>
<td>94.7</td>
</tr>
<tr>
<td>Ohio</td>
<td>90.0</td>
<td>88.3</td>
<td>89.4</td>
</tr>
<tr>
<td>Oklahoma</td>
<td>84.3</td>
<td>86.7</td>
<td>86.0</td>
</tr>
<tr>
<td>Oregon</td>
<td>89.6</td>
<td>82.6</td>
<td>75.4</td>
</tr>
<tr>
<td>Pennsylvania</td>
<td>90.2</td>
<td>89.4</td>
<td>87.6</td>
</tr>
<tr>
<td>Rhode Island</td>
<td>87.9</td>
<td>89.4</td>
<td>86.1</td>
</tr>
<tr>
<td>So. Carolina</td>
<td>85.0</td>
<td>87.8</td>
<td>87.6</td>
</tr>
<tr>
<td>South Dakota</td>
<td>89.1</td>
<td>91.3</td>
<td>89.8</td>
</tr>
<tr>
<td>Tennessee</td>
<td>75.7</td>
<td>84.5</td>
<td>86.9</td>
</tr>
<tr>
<td>Texas</td>
<td>80.0</td>
<td>79.5</td>
<td>80.2</td>
</tr>
<tr>
<td>Utah</td>
<td>93.9</td>
<td>93.4</td>
<td>90.7</td>
</tr>
<tr>
<td>Vermont</td>
<td>87.0</td>
<td>88.1</td>
<td>93.6</td>
</tr>
<tr>
<td>Virginia</td>
<td>88.6</td>
<td>87.5</td>
<td>85.9</td>
</tr>
<tr>
<td>Washington</td>
<td>90.7</td>
<td>85.7</td>
<td>87.7</td>
</tr>
<tr>
<td>West Virginia</td>
<td>83.3</td>
<td>86.8</td>
<td>89.1</td>
</tr>
</tbody>
</table>
### Comparing evidence on dropouts in Texas

We have now described and summarized five different sources of evidence on dropout rates in Texas: 1) dropout data reported by the TEA; 2) IDRA attrition analysis results; 3) the most recent NCES report on high school completion, based on CPS surveys; 4) cohort progression analyses from grade 9 to high school graduation and from 6 to high school graduation discussed in Part 5 above; and 5) estimated dropouts for 1996-97 based on 1995-96 grade enrollments and 1996-97 retention rates (reported in Section 5.5 above). How can we make sense of these vastly different estimates of the extent of the dropout problem in Texas, with dropout rate estimates for the late 1990s ranging from a low of 14.7% reported by the TEA as the "1997-98 actual longitudinal dropout rate" for grades 7 through 12, to a high of the 42% attrition rate reported by IRDA, also for 1997-98, but only for grades 9 through 12?

First, it seems clear that the TEA-reported dropout rates can be largely discounted, as inaccurate and misleading. A November 1999 report from the Texas House Research Organization, *The Dropout Data Debate*, recounts that "In 1996, the State Auditor's Office estimated that the 1994 dropout numbers reported by the Texas Education Agency (TEA) likely covered only half of the actual number of dropouts" (p. 1). The report goes on to recount numerous problems in TEA's approach to calculating dropout rates including changing rules over time in how to define dropouts, relying on district reports of dropouts, while at the same time, beginning in 1992-93 using dropout rate as a key factor in TEA's accountability ratings of districts, and apparent fraud in district reporting. The TEA has developed a system for classifying school leavers in dozens of different ways and many types of "leavers" are not counted as dropouts. Indeed, in 1994, the TEA started classifying students who "met all graduation requirements but failed to pass TAAS" as non-dropout "leavers."

Second, based on a comparison of the cohort progression analyses from grade 9 to high school graduation with those from 6 to high school graduation, it seems clear that the IRDA attrition analyses may represent somewhat inflated estimates of the extent of dropouts because of the increased rate of retention of students in grade 9 (see Figure 5.3). Still the IDRA approach does have one virtue as compared with cohort progression analyses; namely, it attempts to adjust for net immigration of students into Texas schools. I return to this point later. But first let us compare the other three sources of evidence.

The estimates of dropouts for 1996-97 based on 1995-96 grade enrollments and 1996-97 retention rates indicated that about 68,000 high school students dropped out of school between 1995-96 and 1996-97. Adding the missing students across the three grades to estimate longitudinal dropout rates suggests overall dropout rates of 27%.

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Wisconsin</td>
<td>92.4</td>
<td>93.5</td>
<td>90.8</td>
</tr>
<tr>
<td>Wyoming</td>
<td>92.0</td>
<td>90.8</td>
<td>87.6</td>
</tr>
<tr>
<td>Min</td>
<td>75.7</td>
<td>78.7</td>
<td>75.4</td>
</tr>
<tr>
<td>Max</td>
<td>96.3</td>
<td>96.4</td>
<td>94.7</td>
</tr>
<tr>
<td>Mean</td>
<td>87.6</td>
<td>88.1</td>
<td>87.1</td>
</tr>
<tr>
<td>Median</td>
<td>87.9</td>
<td>88.3</td>
<td>87.6</td>
</tr>
</tbody>
</table>

across grades 10-12 (22.5% for White, 33.7% for Black and 38.5% for Hispanic students). These estimates correspond relatively well with the grade 6 to high school graduation cohort analyses (results of which were graphed in Figures 5.5 and 5.6). These results showed that of grade 6 students in the cohort class of 1997, 75.8% of White students and 61.1% of Black and Hispanic students graduated in 1997, implying that 24.2% of White and 38.9% of minority students did not graduate and may have dropped out. Overall for the cohort class of 1997, 31% of the students in grade 6 in 1990-91 did not graduate in 1997. (The 27% figure just cited is slightly lower, presumably because it does note take into account students who drop out between fall of grade 12 and high school graduation the following spring).

Can these results be reconciled with the most recent NCES report on high school completion, based on CPS surveys? Recall that this report indicated that for 18 through 24 year-olds in Texas (not currently enrolled in high school or below) surveyed in October 1996-98, 80.2% reported completing high school. This implies a non-completion or dropout rate of 19.8%. (The CPS survey samples on which this estimate is based are not large enough to derive separate estimates by ethnic group.) It should be noted first that the CPS surveys of 18-24 year-olds in 1996-1998, do not correspond very precisely with the cohort of students in the Texas class of 1997. Nonetheless, two other factors may explain why the CPS derived non-completion (or dropout) estimate of 19.8% is so much lower than 31% estimate derived above for the class of 1997.

One possibility suggested by a previous National Research Council report is that the CPS household surveys tend to under-represent minority youth generally and to underestimate high school dropout rates specifically. In discussing evidence on educational attainment of Black youth, Jaynes and Williams (1989, p. 338) comment that "after age 16, there are very serious, and perhaps growing, problems of surveying the black population, especially black men," and go on to discount a dropout estimate from CPS data from the 1980s for Blacks as simply not credible. If the CPS surveys do in fact under-represent minority youth, this would deflate the overall dropout estimates for Texas derived from this source, since all indications (even those from the TEA) are that dropout rates in Texas are higher for Black and Hispanic youth than for White youth. (Note 20)

The other possibility, alluded to previously, is that the CPS surveys are based on self-reports of high school completion whether it be via normal high school completion or via alternative high school completion such as the GED testing. To explore this possibility, I consulted annual Statistical Reports from the GED Testing Service (1990-1998). Before presenting results from this source, it may be useful to explain the GED Testing program briefly.

The Tests of General Educational Development were developed during World War II to provide adults who did not complete high school with an opportunity to earn a high school equivalency diploma. There are five GED tests: Writing Skills, Social Studies, Science, Interpreting Literature and the Arts, and Mathematics. States and other jurisdictions that contract to use the GED tests establish their own minimum scores for award of the high school equivalency diploma, with the condition that state minimum requirements cannot be below a floor approved by the Commission on Educational Credit and Credentials (an agency of the American Council on Education). For most of the past 10 years, the approved minimum was that examinees had to attain standard scores of at least 40 on each of the five GED tests or an average standard score of at least 45. "In the United States, this minimum standard of 'Minimum 40 or Mean 45' was met by an estimated 75% of the 1987 high school norm group." (GED Testing Service,
In the early 1990s, four states were using this Commission-approved minimum passing standard on the GED tests for award of the high school equivalency degree: Louisiana, Mississippi, Nebraska, and Texas. An additional 27 states were using a similarly low "Minimum 35 and Mean 45" standard. The GED has been widely used in Texas; and in 1996, Texas became the first state in the nation to issue more than 1,000,000 GED credentials since 1971, when the GED started tracking this statistic" (GED Testing Service, 1997, *GED 1996 Statistical Report*, p. 27).

About this time, in keeping with the national movement to raise educational standards, the GED Testing Service decided to raise the minimum passing score on the GED:


Given this background, let us now examine the evidence on GED taking in Texas. Figure 7.2 shows the numbers of people taking and passing the GED (complete battery) from 1989 to 1998. As can be seen, the numbers taking the GED in Texas increased steadily between 1989 and 1996, from about 47,000 to 74,000, a increase of 57% (during the same interval the increase in GED taking nationally was about 26%). GED statistics also make it clear that during this interval the Texas GED-taking population was younger than the national GED-taking population. Over this interval from 25% to 30% of GED takers in Texas were reported to be age 18 or less. (Note 22)

The sharp upturn in GED taking in Texas between 1995 and 1996 (from 74,000 to
87,000, a 17.5% increase) seems readily explained by anticipation of the increase in the GED passing score as of January 1, 1997 (nationally there was a 5% increase in GED test taking between 1995 and 1996). As the GED Testing Service *GED 1997 Statistical Report* explains "The five percent increase in 1996 is most likely attributed to adults attempting to complete the battery before implementation of the 1997 standard" (GED Testing Service, 1998, p. iii).

As a result of the new GED Testing Service minimum passing standard for 1997, 36 jurisdictions were required to raise their passing standard in 1997. Texas was one of them. Surely not coincidentally, the number of people taking the GED in Texas in 1997 dropped from 87,000 to 61,000—an almost 30% decrease. Nationally there was a 5% decrease in GED-taking between 1996 and 1997.

Among the 36 jurisdictions required to increase their passing scores on the GED between 1996 and 1997, "the passing rate decreased by 3.8 percent from 1996 (71.8 percent) to 1997 (68 percent)" (GED Testing Service, 1999, p. 6). In Texas, the GED passing rate fell from 75.2% to 64.2%. This 11% decrease in the passing rate was almost triple the average decrease among the 36 jurisdictions that were required to increase the GED passing scores in 1997. (Note 23)

These developments regarding the GED in Texas suggest a clear explanation for why the percentages of the cohort classes of 1997, 1998 and 1999, began to show slight increases in the percentages of students progressing from grade 6 to high school graduation (for minorities from 60% to 65% and for Whites from 75% to 77%, see Figures 5.6 and 5.7). After the requirements for passing the GED in Texas were stiffened in 1997, and the GED pass rate fell sharply, it appears likely that more students in Texas decided to persist in school to graduation instead of seeking the alternative certification via the more difficult GED standard required by the GEDTS as of January 1, 1997. (Note 24)

Now we can return to the question that prompted my study of GED data. Can GED credentialling in Texas explain why the CPS derived non-completion (or dropout) estimate of 19.8% is so much lower than the 31% non-graduation rate derived from analyses of progress of the cohort class of 1997 from grade 6 to high school graduation? Before addressing this question let me note that neither GED Testing Service data, nor CPS-reported high school completion data are available at the state level disaggregated by ethnicity, so we will have to address this issue across the three major ethnic groups in Texas, namely, White, Black and Hispanic. In 1990-91, according to TEA statistics there were are total of 256,000 White, Black and Hispanic students enrolled in grade 6 in Texas. Eleven per cent (i.e., the difference between the 20% non-completion rate indicated by CPS results and the 31% non-graduation rate derived from the cohort analyses) equals about 28,000. This number—28,000—appears strikingly smaller than the numbers of people who were taking and passing the GED in Texas in 1996 and 1997 (see Figure 7.2). But it must be recalled that though the Texas population of GED takers is younger than the national population of GED takers, only about 35% of GED test takers in 1997 were age 18 or less. If we assume that 35% of the 40,000 GED test-takers in Texas who passed in 1997 might have been members of the cohort class of 1997 (surely a liberal estimate) we get 14,000. This suggests that while GED-taking may account for a substantial portion of the difference between estimates of non-completion of high school based on our cohort analyses (31%) and from CPS-derived estimates (20%), it may not account for all of the difference.

Before summarizing conclusions from this discussion of different sources of evidence on dropout rates in Texas, let me mention briefly two other sources of evidence, and explain why the TEA's exclusion of GED aspirants from its definition of
dropouts is misleading. The first additional source of evidence is from the Annie E. Casey Foundation and in particular, the Casey Foundation's 2000 KIDS Count on-line data base. I was alerted to this source by Hauser (1997), who, while pointing out many limitations of CPS data for estimating dropout rates, also mentions that KIDS Count project as using CPS data in an unusual way to try to obtain relatively current evidence on dropouts across the states. Specifically, this project has compiled from CPS data three-year rolling average estimates from 1985 to 1997 of the percentage of teens ages 16-19 who are dropouts and the percentage of teens not attending school and not working. Since the 2000 KIDS Count results are readily available on-line in table, graph and down loadable database form (www.aecf.org/kidscount/kc1999/), I do not discuss them in detail here. Suffice it to say that: 1) according to both indicators of youth welfare, between 1985 and 1997, Texas had one of the poorer records among the states, consistently showing more than 10% of teens ages 16-19 as dropouts and more than 10% of teens not attending school and not working; and 2) if one examines the standing of Texas on these two indicators relative to those of other states, conditions in Texas seemed to have worsened in the early 1990s after implementation of TAAS.

Second, in a remarkable research effort for MALDEF in the TAAS case, Mark Fassold assembled longitudinal data sets on the Texas sophomore cohorts of 1994 and 1995 (the classes of 1996 and 1997). Using these data sets, Fassold (1999) was able to calculate the cumulative rates of passing the TAAS exit test for up to ten administrations of the test for which students were eligible before their scheduled graduation. He found that the cumulative pass rates for the classes of 1996 and 1997 were 85.2% and 87.1% for White students, 62.3% and 66.1% for Blacks and 65.9% and 69.4% for Hispanics. These results indicate that the White non-graduation rate was in the range of 13-15%, for Blacks 34-38% and Hispanics 30-34%. Fassold's results correspond reasonably well with the cohort progression analyses presented in Part 5 above—especially when two factors are noted. First, Fassold's analysis excluded students classified as special education students. As we saw in part 5.6 above, some 5 to 7% of students taking the TAAS exit test in recent years have been have been classified as special education. Second it is important to note that Fassold's analysis began with grade 10 enrollments, but we have seen that the largest numbers of students drop out between grade 9 and 10. Before leaving this brief summary of Fassold's analyses, it is worth mentioning that despite criticisms by Texas state attorneys, Judge Prado found Fassold's analyses credible and if anything "likely over-estimated the minority pass rate" (Prado, 2000, p. 16).

As mentioned, TEA's reports on dropouts can be largely discounted, as inaccurate and misleading. But one aspect of the TEA approach to defining dropouts deserves commentary. According to the TEA approach to defining dropouts, a student who leaves school to pursue a GED high school equivalency degree in a state approved program is counted as a school "leaver," but not as a dropout. This approach is potentially misleading for a number of reasons. Here I will explain two. First, the common meaning of the term "dropout" is a student who leaves school without graduating from high school. In this sense, students who leave high school without graduating, whether or not they pursue a GED high school equivalency degree, are dropouts. At the same time, there is support for Texas's practice of not counting students enrolled in secondary school programs aimed at preparing for the GED as dropouts in the NCES Common Core of Data definitions (see Winglee et al., 2000, for a recent discussion of the problem of defining dropouts).

Nonetheless, recent research suggests that despite the term "high school equivalency degree," obtaining such certification is not equivalent to normal high school
graduation and moreover, relatively lax standards for GED certification, as in Texas, can encourage students to drop out of high school before graduation. As Chaplin (1999, p. 2) recounts, "Recent evidence . . . suggests that dropping out to get a GED would be a very costly decision (Cameron and Heckman, 1993; Murnane, Willett, and Tyler, 1998)." He goes on to conclude that "the most reliable evidence generally suggests that obtaining a GED instead of a regular high school degree is likely to result in substantially lower earnings later in life." (Chaplin, 1999, p. 6). (Note 25) Indeed, the earning power of GED recipients appears to be more similar to that of dropouts than to high school graduates. Moreover, Chaplin explains:

GED policies which make it easier to get a GED are designed primarily to help high school dropouts. By doing so, however, they may have the perverse effect of encouraging additional students to drop out. This is because by making it easier to get a GED the policies may increase the expected earnings of high school dropouts and, therefore, increase dropout rates. . . . In general less strict GED policies probably increase dropout rates. (Chaplin, 1999, p. 6).

Chaplin presents evidence bearing on this point nationally, but what seems clear is that this is precisely what has happened in Texas through most of the 1990s.

**Conclusions regarding dropouts in Texas.** It is clear that the TEA has been playing a Texas-sized shell game on the matter of counting dropouts. Every source of evidence other than the TEA (including IDRA, NCES, the Casey Foundation's KIDS Count data, Fassold's analyses and my own) shows Texas as having one of the worst dropout rates among the states. (Recall that even the Texas State Auditor's Office estimated that the 1994 dropout numbers reported by the TEA likely covered only half of the actual number of dropouts.) If we adopt the common sense definition that a dropout is a student who leaves school without graduating from high school, analyses of data on enrollment by year, grade and ethnicity (and numbers of high school graduates each year), tell a reasonably clear story of what has happened in Texas over the last two decades. As shown in Figures 5.5 and 5.6, for the cohort classes of 1982 to 1990, the percentage of Black and Hispanic students who progressed from grade 6 to graduation six years later hovered around 65%. For White students, the corresponding percentage started at about 80% and gradually declined to about 75% in 1990. For the cohort class of 1991, the year TAAS was implemented, the percentages fell dramatically, to 55% for minorities and about 68% for White students. Between 1992 and 1996, the corresponding percentages were 60% for minorities and 75% for Whites. Only after Texas was forced by the GED Testing Service to raise its passing standard for receipt of a so-called high school equivalency degree in 1997, did the percentages persisting from grade 6 to high school graduation begin to creep back up, to 65% for minorities in the class of 1999, and for White students to 78% in the same class.

In sum, these results lead me to conclude that since the implementation of the TAAS high school graduation test in 1991, 22-25% of White students and 35-40% of Black and Hispanic students, have not been persisting from grade 6 to regular high school graduation six years later. Overall, during the 1990s the dropout rate in Texas schools was about 30%. As appalling as this result appears, in concluding this discussion of dropout evidence, I should point out that the high school completion and drop out estimates derived from cohort analyses may actually understate the extent of the problem of dropouts (or to use TEA's euphemism, "school-leaving before graduation"). Recall that one of the virtues of the IDRA attrition analyses was that they sought to adjust
estimates for net changes in school populations because of student migration. The results of the cohort progression analyses just summarized implicitly assume that between the ages of 12 (grade 6) and 18 (grade 12), there is no net change in the size of the student population in Texas because of immigration (from either other states or countries). If in fact there is a net out-migration, the dropout estimates just summarized may be too high. If there is a net in-migration into Texas, the estimates are low.

To check on this possibility, I consulted a recent book on the demography of Texas, *The Texas challenge: Population change and the future of Texas* by Murdock et al. (1997). I cannot adequately summarize this interesting book here. Suffice it to say simply that these demographers conclude that between 1990 and 1995, migration into the state of Texas from other states and foreign countries increased relative to what it had been in the 1980s (see Chapter 2). They suggest that annual rates of net migration into Texas have been on the order or 1-2% in the 15 years preceding 1995. The authors do not provide direct estimates of the age distribution of immigrants into Texas, but the overall implication of their results is clear. The estimates of the dropout problem in Texas derived from cohort progression analyses are somewhat low because they fail to take into account net in-migration of school age youth into the schools of Texas. (Note 26) But to be absolutely clear (and to avoid getting into semantic arguments about the meaning of the term "dropout"), I readily acknowledge that what the cohort progression analyses show is the extent of the problem in Texas of students failing to persist in school through to high school graduation—regardless whether it is caused by students having to repeat grade 9, failing to pass the exit level version of TAAS, officially "dropping out," opting out of regular high school programs to enter GED preparation classes, or some combination of these circumstances.

**7.2 Patterns of Grade Retention in the States**

As recounted above, previous research indicates clearly that retention of students in grade, especially beyond the early elementary level, tends to increase the probability that students drop out of high school before graduation. As the recent report from the National Research Council succinctly stated, "In secondary school, grade retention leads to reduced achievement and much higher rates of school dropout" (Heubert & Hauser, 1999, p. 285). For this reason, I sought to analyze rates of grade retention across the states (as reported in Heubert & Hauser, 1999, Table 6.1 corrected) in a variety of ways and to see if there was a relationship between rates of retention at the secondary level and rates of high school completion subsequently reported by Kaufman et al. (1999).

In their Table 6.1, Heubert & Hauser (1999) reported rates of grade retention (specifically percentages of students retained in grade) for 26 states and the District of Columbia in selected states for years for which such data were available (most other states do not collect grade retention data at the state level). As Heubert & Hauser (1999, p. 137) themselves observe, "Retention rates are highly variable across the states." For example, first grade retention rates are reported as varying from 20% to only 1%. Rates of retention in the high school years are reported to vary similarly, from highs of 21-26% to lows of less than 5%. Using the approach described in Section 5.4 above, I have analyzed rates of cumulative promotion and retention. Not surprisingly, cumulative chances of retention also vary widely. For example, in Mississippi and the District of Columbia, in recent years the chance of students being retained in grades 1 through 3 are more than 20%, while in other states (such as Maryland and Arizona) chances are less than 5%.

To explore the possible link between retention in grade 9 and high school
completion, I merged data from Heubert & Hauser's Table 6.1 with data from the recent NCES *Dropouts in the United States 1998* report (Kaufman et al., 1999). The resulting data set is shown in Table 7.3.

### Table 7.3
Grade 9 Retention and High School Completion in the States

<table>
<thead>
<tr>
<th>State</th>
<th>Year</th>
<th>Grade 9 Retention Rate</th>
<th>High school completion rate 18-24 year-olds, 1996-98</th>
</tr>
</thead>
<tbody>
<tr>
<td>Alabama</td>
<td>1996-97</td>
<td>12.6%</td>
<td>84.2%</td>
</tr>
<tr>
<td>Arizona</td>
<td>1996-97</td>
<td>7.0</td>
<td>77.1</td>
</tr>
<tr>
<td>District of Columbia</td>
<td>1996-97</td>
<td>18.7</td>
<td>84.9</td>
</tr>
<tr>
<td>Florida</td>
<td>1996-97</td>
<td>14.3</td>
<td>83.6</td>
</tr>
<tr>
<td>Georgia</td>
<td>1996-97</td>
<td>13.1</td>
<td>84.8</td>
</tr>
<tr>
<td>Kentucky</td>
<td>1995-96</td>
<td>10.7</td>
<td>85.2</td>
</tr>
<tr>
<td>Maryland</td>
<td>1996-97</td>
<td>10.3</td>
<td>94.5</td>
</tr>
<tr>
<td>Massachusetts</td>
<td>1995-96</td>
<td>6.3</td>
<td>90.6</td>
</tr>
<tr>
<td>Michigan</td>
<td>1995-96</td>
<td>4.8</td>
<td>91.0</td>
</tr>
<tr>
<td>Mississippi</td>
<td>1996-97</td>
<td>19.7</td>
<td>82.0</td>
</tr>
<tr>
<td>New York</td>
<td>1996-97</td>
<td>19.5</td>
<td>84.7</td>
</tr>
<tr>
<td>North Carolina</td>
<td>1996-97</td>
<td>15.8</td>
<td>85.2</td>
</tr>
<tr>
<td>Ohio</td>
<td>1996-97</td>
<td>11.4</td>
<td>89.4</td>
</tr>
<tr>
<td>Tennessee</td>
<td>1996-97</td>
<td>13.4</td>
<td>86.9</td>
</tr>
<tr>
<td>Texas</td>
<td>1995-96</td>
<td>17.8</td>
<td>80.2</td>
</tr>
<tr>
<td>Vermont</td>
<td>1996-97</td>
<td>4.8</td>
<td>93.6</td>
</tr>
<tr>
<td>Virginia</td>
<td>1995-96</td>
<td>13.2</td>
<td>85.9</td>
</tr>
<tr>
<td>Wisconsin</td>
<td>1996-97</td>
<td>8.5</td>
<td>90.8</td>
</tr>
</tbody>
</table>

Sources: Heubert & Hauser (1999) Table 6.1; Kaufman et al. (1999), Table 5.

Note that from the first source I took the grade 9 retention rate for 1995-96 or 1996-97, whichever was latest. Note also that the high school completion rates suffer from the problems discussed earlier regarding CPS data as a source of evidence on high school graduation and dropouts. Nonetheless even a casual inspection of these data reveals a clear pattern. States with the higher rates of grade 9 retention tend to have lower rates of high school completion. This pattern can be seen more clearly in Figure 7.3. (Note 27)

Interestingly, Texas with a grade 9 retention rate of 17.8% has a slightly lower high school completion rate (80.2%) than we would expect given the overall pattern among the states shown in Figure 7.3--even though, as previously discussed this rate for
Texas may well be inflated relative to other states because of the high rate of GED taking in Texas. Obviously, such a correlation between two variables, in this case, higher rates of grade 9 retention associated with lower rates of high school completion, does not prove causation, but such a relationship certainly tends to confirm the finding from previous research that grade retention in secondary school leads to higher rates of students dropping out of school before high school graduation.

![Graph showing scatter plot of Grade 9 Retention Rate vs. High School Completion Rate for 18 States](image)

Figure 7.3  Scatter plot of Grade 9 Retention and High School Completion Rates for 18 States

7.3 SAT Scores

It is clear that a substantial portion of the increased pass rates on the TAAS exit test between 1991 and 1998 is, as mentioned previously, an illusion based on exclusion. Specifically, much of the apparent increase in grade 10 TAAS pass rates is due to increased numbers of students taking the grade 10 exit level version of TAAS being classified as special education students, and increased rates of students dropping out of high school in Texas, at least until 1997. When the low standard in Texas for passing the GED had to be raised because the GED Testing Service set a new minimum passing standard as of January 1, 1997, this seems to have had the effect of encouraging a few percentage points more students to persist in school to graduation.

Nonetheless, as best I can estimate, about half of the apparent increase in TAAS exit level pass rates cannot be attributed to such exclusions. So it is relevant to address the question of whether gains on TAAS are a real indication of increased academic learning among students in Texas or whether they represent scores inflated due to extensive preparation for this particular test.

To help answer this question, it is necessary to look at other evidence of student learning in Texas, to see whether the apparent gains on TAAS since its introduction in 1991 are reflected in any other indicators of student learning in Texas. I now summarize evidence from the SAT college admissions test—the test that used to be called the Scholastic Aptitude Test, briefly (and redundantly) the Scholastic Assessment Test, and now is officially named SAT-I.

SAT scores are reported separately for the verbal (SAT-V) and math (SAT-M) portions of this college admissions test, on a scale ranging from 400 to 800 for each
sub-test. Using data from the College Board on SAT scores for the states, I examined performance on the SAT of students in Texas compared to students nationally from a number of perspectives (state rankings on the SAT-V and SAT-M from the 1970s to the 1990s, relative performance of different ethnic groups of students, performance of all SAT-takers vs. high school senior test-takers, etc.). I will not try to summarize results of all of these analyses here. Suffice it to say that the general conclusion of these analyses is that, at least as measured by performance on the SAT, the academic learning of secondary school students in Texas has not improved since the early 1990s, at least as compared with SAT-takers nationally. (Source: College Board, State SAT Scores, 1987-1998, Number of SAT Candidates with Verbal and Math Mean Scores and Standard Deviations—National and for each State, 1972 through 1998, and Report on the Record Numbers of Students in the High School Class. (press release dated August 31, 1999).)

Summary results of two sets of analyses of Texas students' performance on the SAT compared with students nationally from 1972 to 1999 are shown in Figures 7.4 and 7.5. As can be seen from these figures, the performance of Texas students on the SAT was relatively close to the national average in 1970s, but beginning around 1980, increasingly large gaps were apparent on both the SAT-V and SAT-M between national and Texas average scores. There was a slight narrowing in the Texas-national gap on the SAT-M from about 1990 until 1993, but from 1993 to 1998, the gap has increased such that in 1999, on average Texas students were scoring 12 points below the national average on the SAT-M (499 vs. 511).

In short, the pattern of results on both the SAT-V and SAT-M for Texas secondary school students relative to students nationally fails to confirm the gains on the exit level TAAS during the 1990s. Moreover, the pattern of results on the SAT-M indicates that at least since 1993, Texas students' performance on the SAT has worsened relative to students nationally. (Note 28)
One possible explanation for why gains on TAAS do not show up in gains on the SAT is that increasing numbers of students in Texas have been taking the SAT over the last three decades. Not surprisingly, state officials in Texas have advanced this idea to explain the obvious discrepancy between dramatic gains on TAAS in the 1990s and the relatively flat SAT scores for students in Texas. To evaluate this possibility, we can look at numbers of students taking the SAT annually from 1972 to the present. It is indeed true that the numbers of students taking the SAT in Texas have increased faster (from around 50,000 annually through most of the 1970s to 100,000 in 1998) than nationally (from about 1 million annually to 1.2 million recently). However, it is also true that over this period the population of Texas has been increasing far faster than the U.S. population. Murdock et al. (1997, p. 12) report for instance that the population of Texas grew from 11.2 million in 1970 to 18.7 million in 1995 (a 67% increase) compared to a national population increase from 203 million to 263 million (a 29% increase). They also point out that the youth population of Texas in particular has been growing faster than the national youth group.

A better way to evaluate the hypothesis that increases in SAT-taking in Texas explain the flat pattern in SAT scores is to compare the numbers of SAT-takers to the high school population. One such statistic is reported by the College Board, namely the percent of high school graduates taking the SAT. Figure 7.6 shows relevant data for the 50 states for 1999. Specifically, this figure shows state average SAT-M scores for 1999 compared with the percentage of high school graduates in 1999 taking the SAT.
As can be seen in Figure 7.6, there is a clear relationship between these two variables. States with smaller percentages of high school graduates taking the SAT tend to have higher average SAT-M scores. States with larger percentages of high school graduates taking the SAT tend to have lower average SAT-M scores.

What about Texas? According to College Board data, in 1999 Texas had 50% of high school graduates taking the SAT, scoring on average 499 on the SAT-M. This means that Texas, according to the pattern shown in Figure 7.6, has a slightly lower SAT-M average than states with comparable percentages of high school graduates taking the SAT. For example, according to the 1999 College Board data, there were seven states that had between 49% and 53% of high school graduates taking the SAT (Alaska, California, Florida, Hawaii, Oregon, Texas and Washington). Among these states Texas had the lowest SAT-M average in 1999 (499), except for Florida (498). Leaving aside Florida, Texas had an SAT-M average 15-25 points below states with comparable percentages of high school graduates taking the SAT. These results clearly indicate that the relatively poor standing of Texas among the states on SAT scores cannot be attributed to the proportion of secondary school students in Texas taking the SAT.

Moreover, the College Board data may actually understate the relatively poor performance of Texas students on the SAT. This is because Texas has such a poor record regarding student progress to grade 12 and graduation. Even if we use the very conservative estimates of high school completion derived from CPS data (and reproduced in Table 7.2 above) we see that Texas has a rate of non-completion of high school among young adults of about 20%—more than 5 percentage points above the national rate (and as the discussion in Section 7.1 indicated, this figure surely understimates the extent of the high school dropout problem in Texas).

7.4 NAEP Scores Revisited

As mentioned in Section 3.4 above, 1996 NAEP mathematics scores were released in 1997 and seemed to provide confirmation that gains apparent on TAAS were
<table>
<thead>
<tr>
<th>Grade 4</th>
<th>1994 Mean (SS)</th>
<th>1995 Mean (SS)</th>
<th>1996 Mean (SS)</th>
<th>1997 Mean (SS)</th>
<th>1998 Mean (SS)</th>
<th>1999 Mean (SS)</th>
<th>Gain 1994-99</th>
<th>Gain/SD (SD for TLI=15; SD for x-level writing test est'd = 200)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Reading</td>
<td>4-78.4</td>
<td>4-80.1</td>
<td>4-79.9</td>
<td>4-80.9</td>
<td>4-84.4</td>
<td>4-85.3</td>
<td>6.6</td>
<td>0.44</td>
</tr>
<tr>
<td>Math</td>
<td>4-70.5</td>
<td>4-74.6</td>
<td>4-77.3</td>
<td>4-79.0</td>
<td>4-80.0</td>
<td>4-80.9</td>
<td>10.4</td>
<td>0.69</td>
</tr>
<tr>
<td>Writing</td>
<td>1640</td>
<td>1647</td>
<td>1646</td>
<td>1663</td>
<td>1670</td>
<td>1673</td>
<td>33</td>
<td>--</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Grade 8</th>
<th>1994 Mean (SS)</th>
<th>1995 Mean (SS)</th>
<th>1996 Mean (SS)</th>
<th>1997 Mean (SS)</th>
<th>1998 Mean (SS)</th>
<th>1999 Mean (SS)</th>
<th>Gain 1994-99</th>
<th>Gain/SD (SD for TLI=15; SD for x-level writing test est'd = 200)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Reading</td>
<td>8-77.8</td>
<td>8-78.0</td>
<td>8-79.8</td>
<td>8-81.8</td>
<td>8-83.3</td>
<td>8-84.7</td>
<td>6.9</td>
<td>0.46</td>
</tr>
<tr>
<td>Math</td>
<td>8-70.0</td>
<td>8-69.7</td>
<td>8-73.8</td>
<td>8-76.7</td>
<td>8-78.7</td>
<td>8-80.8</td>
<td>10.8</td>
<td>0.72</td>
</tr>
<tr>
<td>Writing</td>
<td>1591</td>
<td>1606</td>
<td>1611</td>
<td>1631</td>
<td>1655</td>
<td>1663</td>
<td>72</td>
<td>--</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Grade 10</th>
<th>1994 Mean (SS)</th>
<th>1995 Mean (SS)</th>
<th>1996 Mean (SS)</th>
<th>1997 Mean (SS)</th>
<th>1998 Mean (SS)</th>
<th>1999 Mean (SS)</th>
<th>Gain 1994-99</th>
<th>Gain/SD (SD for TLI=15; SD for x-level writing test est'd = 200)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Reading</td>
<td>x-77.7</td>
<td>x-77.8</td>
<td>x-80.0</td>
<td>x-82.1</td>
<td>x-83.9</td>
<td>x-84.8</td>
<td>7.1</td>
<td>0.47</td>
</tr>
<tr>
<td>Math</td>
<td>x-69.9</td>
<td>x-71.2</td>
<td>x-72.9</td>
<td>x-75.2</td>
<td>x-77.4</td>
<td>x-79.3</td>
<td>9.4</td>
<td>0.63</td>
</tr>
<tr>
<td>Writing</td>
<td>1648</td>
<td>1677</td>
<td>1685</td>
<td>1719</td>
<td>1708</td>
<td>1734</td>
<td>86</td>
<td>0.43</td>
</tr>
</tbody>
</table>

Source: www.tea.tx.state.us/student.assessment/results/summary/

The last column in Table 7.4 shows the 1994 to 1999 gains on TAAS divided by the relevant TAAS test standard deviation (15 for the reading and math TAAS tests and 200 for the TAAS exit level writing test). These results, average test score changes divided by the relevant standard deviations, may be interpreted as effect sizes.

Before discussing the meaning of the results shown in the last column of Table 7.4, a brief summary of the idea of effect size may be helpful. (Yes, dear reader, yet another digression. But if you know about meta-analysis and effect size, just skip ahead.) The concept of effect size has come to be widely recognized in educational research in the last two decades because of the increasing prominence of meta-analysis. Meta-analysis refers to the statistical analysis of the findings of previous empirical studies. With the proliferation of research studies on particular issues, statistical analysis and summary of patterns across many studies on the same issue have proven to be a useful tool in understanding patterns of findings on a research issue (Glass, 1976; Cohen, 1977; Glass, McGaw & Smith, 1981; Wolf, 1986; Hunter & Schmidt 1990, and Cooper & Hedges, 1994 are some of the basic reference works on meta-analysis). In meta-analysis, effect size is defined as the difference between two group mean scores expressed in standard
score form, or—since the technique is generally applied to experimental or quasi-experimental studies—the difference between the mean of the treatment group and the mean of the control group, divided by the standard deviation of the control group (Glass, McGaw & Smith, 1981, p. 29). Mathematically this is generally expressed as:

\[
\text{Effect size} = \frac{\bar{X}_t - \bar{X}_c}{s_c}
\]

where \(\bar{X}_t\) is the treatment mean, 
\(\bar{X}_c\) is the control mean, and
\(s_c\) is the control group standard deviation.

Interpretation of magnitude of effect sizes varies somewhat according to different authorities, but one commonly cited rule of thumb is that an effect size of 0.2 constitutes a small effect, 0.5 a medium effect and 0.8 a large effect (Cohen, 1977, Wolf, 1986, p. 27). As a general guideline, the Joint Dissemination Review Panel of the National Institute of Education adopted the approach that an effect size had to be one-third (0.33) or at least one-quarter (0.25) of a standard deviation in order to be educationally meaningful (Wolf, 1986, p. 27).

While meta-analysis has been applied in many areas of social science research, perhaps most directly relevant to interpretation of TAAS and NAEP score changes are studies which have employed meta-analysis to examine the effects of test preparation and coaching. Becker's (1990) analysis of previous studies of the effectiveness of coaching for the SAT is a good example of such a study. Though she used a metric for comparing study outcomes which is somewhat unusual in the meta-analysis literature—namely the standardized mean-change measure—this measure is computed in standard deviation units, and is directly analogous to effect size. Becker analyzed study outcomes in terms of some 20 study characteristics having to do with both study design and content of coaching studied. Like previous analysts, she found that coaching effects were larger for the SAT-M than for the SAT-V. However, unlike some previous researchers, she did not find that duration of coaching was a strong predictor of the effects of coaching. Instead, she found that of all the coaching content variables she investigated, "item practice," (i.e., coaching in which participants were given practice on sample test items, was the strongest influence on coaching outcomes). Overall, she concluded that among 21 published comparison studies, the effects of coaching were 0.09 standard deviations of the SAT-V and 0.16 on SAT-M.

Against this backdrop, the gains on TAAS summarized in Table 7.4 appear quite impressive. Across all three grades and all three TAAS subject areas (reading, math and writing), the magnitude of TAAS increases ranged from 0.43 to 0.72 standard deviation units. According to guidelines for interpreting effect sizes, these gains clearly fall into the range of medium to large effects. Also, the gains on TAAS clearly exceed the gains that appear possible, according to previous research, from mere test coaching. (In one respect though, the TAAS gains do parallel results from Becker's study of test coaching: gains on math tests are larger than those on reading tests.) The gains on TAAS seem especially impressive when it is recalled that the gains on TAAS summarized in Table 7.4 represent performance of hundreds of thousand of Texas students, while most of the studies examined via meta-analysis involved mere hundreds or thousands of subjects.

Having re-examined TAAS score changes in Texas from the effect size perspective, we may now turn to revisit NAEP scores for Texas. The fundamental question we address is whether NAEP results for Texas provide confirmation of the dramatic gains apparent on the TAAS. We first consider NAEP results for Texas, overall,
for grade 4 and 8 students and then take a closer look at results disaggregated by ethnic group.

**Table 7.5**  
Mean NAEP Scores, Texas and Nation, Grade 4 and 8, 1990-98

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Mean</td>
<td>SD</td>
<td>Mean</td>
<td>SD</td>
<td>Mean</td>
</tr>
<tr>
<td><strong>Reading, Grade 4</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Texas</td>
<td>213</td>
<td>34</td>
<td>212</td>
<td>39</td>
<td>217</td>
</tr>
<tr>
<td>Nation</td>
<td>216.7</td>
<td>36</td>
<td>214.3</td>
<td>41</td>
<td>217</td>
</tr>
<tr>
<td><strong>Reading, Grade 8</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Texas</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Nation</td>
<td>260.0</td>
<td>36</td>
<td>259.6</td>
<td>37</td>
<td>264.0</td>
</tr>
<tr>
<td><strong>Writing, Grade 8</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Texas</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Nation</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Mathematics, Grade 4</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Texas</td>
<td>217.9</td>
<td>30.3</td>
<td></td>
<td></td>
<td>228.7</td>
</tr>
<tr>
<td>Nation</td>
<td>213.1</td>
<td>31.8</td>
<td>219.7</td>
<td>31.7</td>
<td>223.9</td>
</tr>
<tr>
<td><strong>Mathematics, Grade 8</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Texas</td>
<td>258.2</td>
<td>35.4</td>
<td>264.6</td>
<td>36.8</td>
<td>270.2</td>
</tr>
<tr>
<td>Nation</td>
<td>262.6</td>
<td>36</td>
<td>268.4</td>
<td>36.3</td>
<td>272.0</td>
</tr>
<tr>
<td><strong>Science, Grade 8</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Texas</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Nation</td>
<td>148.5</td>
<td>34.1</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>


There are two perspectives from which to consider the NAEP results for Texas shown here. We may compare the mean scores of Texas 4th and 8th graders with 4th and 8th graders nationally, or for NAEP reading and math state assessments (the only ones done in more than one year), we may look at how the performance of Texas students seems to have changed over time. From the former perspective, it is clear that the performance of Texas 4th and 8th graders is very similar to the performance of 4th and 8th graders nationally. In all eleven instances in which state NAEP assessments allow comparison of student performance in Texas with student performance nationally, there is not a single instance in which average NAEP scores in Texas vary from national means by as much as two-tenths of a standard deviation. Texas grade 8 students scored better than students nationally on the NAEP writing assessment in 1998, but they scored worse on the science assessment in 1996, by about the same amount (+ 0.10 standard deviation units in writing and 0.10 in science). It may be recalled that according to guidelines in the meta-analysis literature, differences of less than one-quarter of a standard deviation are small and not considered educationally meaningful. In reading, at grade 4 we have three
years in which we can compare Texas NAEP reading scores with the national average, 1992, 1994 and 1998. There appears to be a very slight trend for Texas grade 4 reading scores to have converged with the national average between 1992 and 1998; but note, that to begin with, in 1992 the Texas average was only one-tenth of a standard deviation below the national average: \( \frac{216.7-213}{36} = 0.102 \). In grade 8 reading we have a Texas-national comparison for just one year, 1998. In 1998, Texas eighth graders scored on average only very slightly below the national average, but again, the difference was less than one-tenth of a standard deviation: \( \frac{264-262}{35} = 0.057 \).

We also have three years in which we can compare national and Texas NAEP math scores, 1990, 1992 and 1996. In 1992, the Texas NAEP math score average at grade 4 (217.9) was only slightly below the national average (219.7), but by 1996, it was slightly above the national average, by an amount equivalent to about 15% of a standard deviation: \( \frac{228.7-223.8}{31.2} = 0.154 \). For 1990, 1992 and 1996, the Texas NAEP math grade 8 average was slightly below the national average by amounts equivalent to 12%, 10% and 5% of the national standard deviation.

Now, let us put aside national NAEP results and simply consider the gains apparent in state NAEP results for Texas. Between 1994 and 1998, the Texas NAEP reading average increased from 212 to 217, an amount equivalent to 12% of the 1994 national standard deviation \( \frac{5}{41} = 0.122 \). At grade 8, the Texas NAEP math average increased 12 points between 1990 and 1996, an amount equivalent to 33% of a standard deviation \( \frac{12}{36} = 0.33 \). According to the guidelines cited earlier from the meta-analysis literature, this is an amount that qualifies as a small, but educationally meaningful difference.

More germane to the question whether TAAS gains are real is consideration of the magnitude of the gains apparent on TAAS (shown in Table 7.4 above) and those apparent on state NAEP results (shown in Table 7.5). In general, the gains on TAAS, between 1994 and 1999 (in the range of 0.43 to 0.72 standard deviation units) are far larger than the range of gains apparent on NAEP (in the range of 0.12 to 0.33). Unfortunately, there is only one pair of years in which we have results from state NAEP and TAAS for the same subject, namely reading. Between 1994 and 1998, the average grade 4 TLI increased from 78.4 to 84.4, equivalent to 0.40 standard deviations. Between 1994 and 1998, the average grade 4 Texas NAEP score increased from 212 to 217, equivalent to 0.12 standard deviations \( \frac{5}{41} = 0.122 \), and even if we divide by the Texas standard deviation, we get just \( \frac{5}{39} = 0.128 \). Even before we look beneath the surface of NAEP averages for Texas, these results, with gains on NAEP far less than half the size of gains apparent on TAAS (and in the single instance when a direct comparison was possible, NAEP gains of 0.12 were just 30% the size of the 0.40 gain apparent on grade 4 TAAS), suggest clearly that the bulk (at least two-thirds) of the dramatic gains on TAAS are simply not real.

Next, let us delve below the surface of the Texas state NAEP averages and consider the Texas NAEP reading and math averages separately for White, Black and Hispanic students. Table 7.6 shows relevant results for state NAEP reading and math tests.

### Table 7.6

**Texas Mean NAEP Scores by Ethnicity**

**Grade 4 and 8, 1992, 1994 and 1998**

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
As can be seen here, the gap between the average NAEP scores of White students in Texas and those of Black and Hispanic students is fairly consistently in the range of 25 to 35 points. There is a tendency for Hispanic students in Texas to score slightly better on NAEP tests than Black students; but overall, Hispanic and Black students in Texas score on average between two-thirds and a full standard deviation below the mean of White students. Moreover, at grade 4, there is an increase in the White-minority gap in NAEP reading scores between 1992 and 1998. In 1992, the NAEP grade 4 reading average was 224 for White students, 200 for Black students and 201 for Hispanics. By 1998, the corresponding averages were 233, 197 and 204.

Apparent gains for Texas in NAEP math scores between 1992 and 1996 were indeed statistically significant. And in terms of practical significance, critical readers may well be asking themselves, even if the gains were not large in terms of the standard deviation units perspective suggested in the meta-analysis literature, gains on the order of a third of standard deviation, when apparent for a population of a quarter million students (roughly the number of fourth graders in Texas in 1996), are surely are of practical significance. Also, it may be recalled from Section 3.4 above that the NAEP math gains for Texas fourth graders between 1992 and 1996 were greater than the corresponding

<table>
<thead>
<tr>
<th>Reading, Grade 4</th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>White</td>
<td>224</td>
<td>227</td>
<td>232</td>
</tr>
<tr>
<td>Black</td>
<td>200</td>
<td>191</td>
<td>197</td>
</tr>
<tr>
<td>Hispanic</td>
<td>201</td>
<td>198</td>
<td>204</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Reading, Grade 8</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>White</td>
<td>273</td>
</tr>
<tr>
<td>Black</td>
<td>245</td>
</tr>
<tr>
<td>Hispanic</td>
<td>252</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Mathematics, Grade 4</th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>White</td>
<td>228</td>
<td>242</td>
<td></td>
</tr>
<tr>
<td>Black</td>
<td>197</td>
<td>212</td>
<td></td>
</tr>
<tr>
<td>Hispanic</td>
<td>207</td>
<td>216</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Mathematics, Grade 8</th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>White</td>
<td>273</td>
<td>279</td>
<td>285</td>
</tr>
<tr>
<td>Black</td>
<td>236</td>
<td>243</td>
<td>249</td>
</tr>
<tr>
<td>Hispanic</td>
<td>245</td>
<td>248</td>
<td>256</td>
</tr>
</tbody>
</table>

Source: NAEP Data Almanac nces.ed.gov/nationsreportcard/TABLES/index.shtml, Reese et al., 1997; Mullis et al., 1993.
gains for any other state participating in these two NAEP state assessments. So any reasonable person must concede that the apparent improvement of Texas grade 4 NAEP math average from 217.9 in 1992 to 228.7 in 1996 (a gain of about one-third of a standard deviation), if real, is indeed a noteworthy and educationally significant accomplishment.

But there is that "if." The other perspective not yet brought to bear in considering changes in NAEP test score averages is advice offered in Part 1. When considering average test scores, it is always helpful to pay attention to who is and is not tested. NAEP seeks to estimate the level of learning of students in the states not by testing all students in the states in a particular grade, but through use of systematic and representative sample of schools and students. Without getting into details of NAEP sampling, let us focus here on the fact that not all students sampled are actually tested. Some students selected for NAEP testing are excluded because they are limited English proficient (LEP) or because of their status as special education students, whose individualized education plans (IEPs) may call for them to be excluded from standardized testing.

NAEP researchers have long recognized that exclusion of sampled students from NAEP testing has the potential to create bias in NAEP results. Here is how one NAEP report discussed the issue:

The interpretation of comparisons of achievement between two or more assessments depends on the comparability of the populations assessed at each point in time. For example, even if the proficiency distribution of the entire population at time 2 was unchanged from that at time 1, an increase in the rate of exclusion would produce an apparent gain in the reported proficiencies between the two time points if the excluded students tend to be lower performers. (Mullis et al., 1993, p. 353).

Because excluding sampled students from NAEP testing has the potential for skewing results, over time NAEP has developed detailed guidelines for excluding students from testing, and has taken special steps to try to include LEP and special education students in NAEP testing, for example, by allowing accommodations to standard NAEP testing procedures to meet the needs of special education students. (See Reese et al., 1997, Chapter 4 for a discussion of efforts to make NAEP math assessments more inclusive.)

### Table 7.7

**Percentages of IEP and LEP Students Excluded from NAEP State Math Assessments, Texas and Nation**

<table>
<thead>
<tr>
<th></th>
<th>Mathematics, Grade 4</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>1990</td>
<td>1992</td>
<td>1996</td>
</tr>
<tr>
<td>Texas</td>
<td>8%</td>
<td>11%</td>
<td></td>
</tr>
<tr>
<td>Nation</td>
<td>8%</td>
<td>6%</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>Mathematics, Grade 8</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>1990</td>
<td>1992</td>
<td>1996</td>
</tr>
<tr>
<td>Texas</td>
<td>7%</td>
<td>7%</td>
<td>8%</td>
</tr>
<tr>
<td>Nation</td>
<td>6%</td>
<td>7%</td>
<td>5%</td>
</tr>
</tbody>
</table>
Given this background, let us now consider the percentages of students sampled in state NAEP math assessments who were excluded from testing. Table 7.7 shows the percentages of sampled students excluded from testing in NAEP state math assessments in 1990, 1992 and 1996 for both Texas and the nation; recall that in the original trial state NAEP assessment in 1990 only grade 8 was tested. As can be seen in this table, at the national level, between 1992 and 1996, the percentages of students excluded fell slightly (from 8% to 6% at grade 4, and from 7% to 5% at grade 8). These results at the national level were presumably a result of efforts to make NAEP more inclusive in testing LEP and special education students. However, in Texas, the percentages of students excluded from testing increased at both grade levels: from 8% to 11% at grade 4, and from 7% to 8% at grade 8. This means that some portion of the increased NAEP math averages for Texas in 1996 are illusory, resulting from the increased rates of exclusion of LEP and special students in Texas from NAEP testing. The gaps in rates of exclusion between Texas and the nation in 1996 also mean that comparisons of Texas with national averages in that year will be skewed in favor of Texas for the simple reason that more students in Texas were excluded from testing. In short, as with TAAS results, some portion of the apparent gains on NAEP math tests in Texas in the 1990s is an illusion arising from exclusion.

As with TAAS gains, can we estimate what portion of the apparent NAEP gains are real and what portion are artifactual attributable to the increased rates of exclusion of Texas students from NAEP testing? Fortunately, regarding NAEP we have a clear model for estimating the impact of exclusion on NAEP scores. In *NAEP 1992 Mathematics Report card for the Nation and the States*, Mullis et al. (1993, pp. 352-355) discuss the problem of excluding students from NAEP testing and apply a model for estimating the effects of exclusion on distributions of NAEP scores. What these researchers did was to recompute national NAEP results based on the assumption that "all excluded and all absent students, had they been assessed, would have scored below the 25th percentile of all students" (Mullis et al.,1993, pp. 353). Using this approach, we can recompute the NAEP math averages for Texas in 1996, assuming that the percentages of Texas students excluded from NAEP testing were at the national average (6% at grade 4 and 5% at grade 8, as opposed to the observed 11% and 8% exclusions reported for Texas in 1996.).

The NAEP data almanac reports that on the 1996 NAEP math assessments, the scores equivalent to the 10th percentile in Texas were 190.4 and 225.5 for grade 4 and 8, respectively. Using these figures, assuming that the 1996 exclusion rates in Texas were the same as the national rates (and that excluded students in Texas would have scored at the 10th percentile), we may recompute the average grade 4 and grade 8 NAEP math scores for Texas as follows:

\[
\text{Grade 4: } \quad 0.95(228.7) + 0.05(190.4) = 226.9
\]

\[
\text{Grade 8: } \quad 0.97(270.2) + 0.03(225.5) = 268.9
\]

These results indicate that on the order of 20%-25% of the NAEP gains for Texas between 1992 and 1996 were due simply to the high rate of exclusion of students from NAEP testing in 1996. In other words, given these calculations to adjust for the high rates of exclusion of Texas students from NAEP testing in 1996, the gain of scores in Texas from 1992 to 1996 would be 9 points at grade 4 and 4.3 points at grade 8. The former is still considerably above the national increase of 4 points at grade 4, but no longer highest
among the states (North Carolina showed a grade 4 NAEP math gain of 11 points between 1992 and 1996, while excluding just 7% of grade 4 students from testing in 1996). And the gain of 4.3 points at grade 8 would leave Texas very near the level of the national gain apparent between 1992 and 1996.

In summary, review of results of NAEP from the 1990s suggests that grade 4 and grade 8 students in Texas performed much like students nationally. On some NAEP assessments, Texas students scored above the national average, and on some below. In the two subject areas in which state NAEP assessments were conducted more than once during the 1990s, there is evidence of modest progress by students in Texas; but it is much like the progress evident for students nationally. Reviewing NAEP results for Texas by ethnic group, we see a more mixed picture. In many comparisons, Black and Hispanic students show about the same gain in NAEP scores as White students, but the 1998 NAEP reading results, suggest that while White grade 4 reading scores in Texas have improved since 1992, those of Black and Hispanic students have not. More generally, however, the magnitudes of the gains apparent on NAEP for Texas fail to confirm the dramatic gains apparent on TAAS. Gains on NAEP in Texas are consistently much less than half the size (in standard deviation units) of Texas gains on state NAEP assessments. These results indicates that the dramatic gains on TAAS during the 1990s are more illusory than real. The Texas "miracle" is more myth than real.

Before leaving this review of state NAEP results for Texas, it may be helpful to mention Rodamar's (2000) excellent review once more. As mentioned previously, he reviewed TAAS and NAEP results for Texas not in terms of changes measured in standard deviation units, but in terms of percent passing TAAS and percent meeting the NAEP "basic" proficiency standard. While he focused on reading and math test scores (i.e., he did not review NAEP science and writing results), Rodamar reached conclusions very similar to those derived from reviewing NAEP results in terms of effect size changes:

When it comes to educational achievement, by nearly any measure except TAAS, Texas looks a lot like America. Texas was near the national average on many measures of educational performance when TAAS was introduced—and remains there. (Rodamar, 2000, p. 27).

7.5 Other Evidence

TAAS scores, graduation rates, SAT scores, and evidence from NAEP are the most obvious sources of evidence regarding education in Texas. But I have also searched for other evidence that might be available. For example, in its annual review of the "state of the states," Education Week has assembled a wide range of data on a number of dimensions of education in the states (Jerald, 2000). Since this source is widely available, I will not review it in detail. But three findings are worth mentioning. First, Texas received a grade of D in the category of Improving Teacher Quality. Second, the Lone Star state received only middling marks on dimensions of School Climate (C), Resource Adequacy (C+), and Equity (C). Finally, I was struck by the relatively low rate of going to college in Texas. In Texas, in 1996, only 54% of high school graduates were reported to be enrolling in a two- or four-year college, as compared with 65% nationally (Jerald, 2000, p. 71). (Note 30)

This led me to inquire further into another Texas testing program—the Texas Academic Skills Program of TASP test. The Texas Higher Education Coordinating Board describes the TASP testing program thus at its website (www.thecb.state.tx.us/):
You Ready for College?
Are you ready for college courses? Not sure?

Texas, you can find out if you have the reading, writing, and math skills you need to do college-level work through the Texas Academic Skills Program—or TASP. The TASP Test, which is part of the TASP program, is required—it is not optional.

Beginning in fall 1998, you must take the TASP Test, or an alternative test, before beginning classes at a public community college, public technical college, or public university in Texas.

TASP Test is not an admissions test, however. You cannot be denied admission to a public institution of higher education based on your TASP Test score. If you need to improve your skills, you are not alone. About one-half of students entering college need some help. Take the TASP Test while you are in high school so you can identify the skills you need to improve. You'll be confident that you are ready for college.

What have been the results of this "college readiness" testing program? I found the graph reproduced in Figure 7.7 in a report available on the same website.

I could not find more recent results of TASP testing on the Texas Higher Education Coordinating Board website, but Chris Patterson on the Lone Star Foundation of Austin, TX (personal communication March 22, 2000) generously sent me a summary of TASP results from 1993 to 1997, reproduced in Table 7.8 below.

**Table 7.8**  
Annual Texas Academic Skills Program Report  
of Student Performance Pass Rates  
by Race/Ethnicity and Test Area  
1993-1997 High School Graduating Classes

<table>
<thead>
<tr>
<th>Year</th>
<th>Total Count</th>
<th>All 3 Parts Pass Rate</th>
<th>Reading Pass Rate</th>
<th>Math Pass Rate</th>
<th>Writing Pass Rate</th>
</tr>
</thead>
<tbody>
<tr>
<td>1993</td>
<td>64,662</td>
<td>78.0%</td>
<td>90.3%</td>
<td>86.0%</td>
<td>90.3%</td>
</tr>
<tr>
<td>1994</td>
<td>63,257</td>
<td>65.2%</td>
<td>83.2%</td>
<td>79.3%</td>
<td>82.5%</td>
</tr>
<tr>
<td>1995</td>
<td>73,207</td>
<td>51.7%</td>
<td>75.3%</td>
<td>64.3%</td>
<td>80.8%</td>
</tr>
<tr>
<td>1996</td>
<td>68,810</td>
<td>48.1%</td>
<td>74.4%</td>
<td>60.6%</td>
<td>80.0%</td>
</tr>
<tr>
<td>1997</td>
<td>67,833</td>
<td>43.3%</td>
<td>70.7%</td>
<td>55.9%</td>
<td>79.3%</td>
</tr>
</tbody>
</table>

**Native American**

<table>
<thead>
<tr>
<th>Year</th>
<th>Total Count</th>
<th>All 3 Parts Pass Rate</th>
<th>Reading Pass Rate</th>
<th>Math Pass Rate</th>
<th>Writing Pass Rate</th>
</tr>
</thead>
<tbody>
<tr>
<td>1993</td>
<td>107</td>
<td>83.2%</td>
<td>92.5%</td>
<td>89.7%</td>
<td>90.7%</td>
</tr>
<tr>
<td>1994</td>
<td>108</td>
<td>64.8%</td>
<td>82.4%</td>
<td>84.3%</td>
<td>81.5%</td>
</tr>
<tr>
<td>1995</td>
<td>161</td>
<td>52.8%</td>
<td>77.0%</td>
<td>66.5%</td>
<td>89.4%</td>
</tr>
<tr>
<td>1996</td>
<td>136</td>
<td>57.5%</td>
<td>79.4%</td>
<td>69.1%</td>
<td>84.6%</td>
</tr>
<tr>
<td>1997</td>
<td>130</td>
<td>42.3%</td>
<td>64.6%</td>
<td>65.4%</td>
<td>82.3%</td>
</tr>
</tbody>
</table>

**Asian**

<table>
<thead>
<tr>
<th>Year</th>
<th>Total Count</th>
<th>All 3 Parts Pass Rate</th>
<th>Reading Pass Rate</th>
<th>Math Pass Rate</th>
<th>Writing Pass Rate</th>
</tr>
</thead>
<tbody>
<tr>
<td>1993</td>
<td>2,424</td>
<td>79.5%</td>
<td>90.5%</td>
<td>95.7%</td>
<td>84.8%</td>
</tr>
<tr>
<td>1994</td>
<td>2,625</td>
<td>63.0%</td>
<td>78.6%</td>
<td>92.5%</td>
<td>69.3%</td>
</tr>
<tr>
<td>1995</td>
<td>3,168</td>
<td>53.9%</td>
<td>69.7%</td>
<td>85.2%</td>
<td>66.9%</td>
</tr>
<tr>
<td>1996</td>
<td>2,608</td>
<td>49.2%</td>
<td>68.9%</td>
<td>80.8%</td>
<td>66.2%</td>
</tr>
<tr>
<td>1997</td>
<td>2,392</td>
<td>48.5%</td>
<td>66.1%</td>
<td>78.6%</td>
<td>68.7%</td>
</tr>
</tbody>
</table>

**Black**

<table>
<thead>
<tr>
<th>Year</th>
<th>Total Count</th>
<th>All 3 Parts Pass Rate</th>
<th>Reading Pass Rate</th>
<th>Math Pass Rate</th>
<th>Writing Pass Rate</th>
</tr>
</thead>
<tbody>
<tr>
<td>1993</td>
<td>5,678</td>
<td>57.7%</td>
<td>79.8%</td>
<td>69.0%</td>
<td>79.6%</td>
</tr>
<tr>
<td>1994</td>
<td>5,859</td>
<td>44.2%</td>
<td>70.9%</td>
<td>60.3%</td>
<td>69.3%</td>
</tr>
<tr>
<td>1995</td>
<td>7,015</td>
<td>31.2%</td>
<td>60.8%</td>
<td>43.4%</td>
<td>69.3%</td>
</tr>
<tr>
<td>Year</td>
<td>Hispanic</td>
<td>White</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>------</td>
<td>----------</td>
<td>-------</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Students</td>
<td>67.6%</td>
<td>84.2%</td>
<td>84.2%</td>
<td>79.6%</td>
</tr>
<tr>
<td>1993</td>
<td>14,349</td>
<td>84.2%</td>
<td>93.7%</td>
<td>89.9%</td>
<td>93.9%</td>
</tr>
<tr>
<td>1994</td>
<td>15,075</td>
<td>73.1%</td>
<td>88.2%</td>
<td>84.5%</td>
<td>87.9%</td>
</tr>
<tr>
<td>1995</td>
<td>18,121</td>
<td>60.3%</td>
<td>82.0%</td>
<td>70.7%</td>
<td>87.1%</td>
</tr>
<tr>
<td>1996</td>
<td>17,926</td>
<td>57.0%</td>
<td>81.1%</td>
<td>67.6%</td>
<td>86.4%</td>
</tr>
<tr>
<td>1997</td>
<td>19,166</td>
<td>53.0%</td>
<td>78.1%</td>
<td>64.0%</td>
<td>86.4%</td>
</tr>
</tbody>
</table>

Source: Texas Higher Education Coordinating Board
Note: These results reflect pass rates on the initial attempt on the TASP test only.

Reviewing these results from TASP testing, and comparing them with results of TAAS testing (see Figure 3.1 for example), the conclusion seems inescapable that something is seriously amiss in the Texas system of education, the TAAS testing program or the TASP testing program—or perhaps all three. Between 1994 and 1997, TAAS results showed a 20% increase in the percentage of students passing all three exit level TAAS tests (reading, writing and math). But during the same interval, TASP results showed a sharp decrease (from 65.2% to 43.3%) in the percentage of students passing all three parts (reading, math, and writing) of the TASP college readiness test.
8. Summary and Lessons from the Myth Deflated

Before recapping the territory covered in this article and suggesting some of the broader lessons that might be gleaned from the myth of the Texas miracle in education, I pause for one more digression (readers who have made it this far likely will not be too surprised by yet another detour). The detour is to recount a small survey of scholars undertaken in the summer of 1999. After this side excursion, I summarize "the myth of the Texas miracle." Finally, in closing, I suggest some of the broader lessons that might be gleaned from this examination of the illusory Texas miracle.

8.1 The "Two Questions Survey" on School Reform

In August 1999, as I was preparing for the start of the TAAS trial in September, I re-read a number of key documents regarding the development of the TAAS testing program in Texas. One was the Minutes of the Texas State Board of Education in July 1990 (a full copy of these minutes is reproduced in appendix 8 of this article for ease of reference). It may be recalled that it was at this meeting that the Board set the passing scores on TAAS. When reviewing minutes of this meeting, I was struck by the following passage:

Commissioner [of Education in Texas] Kirby reiterated some of the information presented to Committee of the Whole during the Thursday, July 12, 1990, work session on the TAAS, noting the recommendations of the staff regarding this item.

Mr. Davis asked for the rationale for the two-year phase in rather than going immediately to the 70% [passing score on TAAS] or a one-year phase in. The commissioner stated that this would give the board an opportunity to clearly set that 70% is the standard--to state the expectation and expect the schools to present the skills to the students and help the students develop those skills so that this is not an unreasonable expectation. Dr. Kirby said that since this is a different, more difficult test, the needed phase-in time is suggested at least until the results of the fall administration are known. Mr. Davis expressed concern that the test does not appear to be indicative of what is being presented in the classroom. Commissioner Kirby replied that the test is an accurate measurement of what students should be learning, but the test is moving much further in the areas of problem solving, higher order thinking skills, making inferences, and drawing conclusions. He said that it is not believed that at this point in time every student has been adequately prepared in those skills, because with the Texas Educational Assessment of Minimum Skills (TEAMS) tests, emphasis has been placed on the basic skills. The commissioner noted that the test drives the curriculum and that it
will require a year or two to make that kind of adjustment in the focus of the curriculum. (TEA, 1997, Appendix 9 of the *Texas Student Assessment Program Technical Digest for the Academic Year 1996-1997*, pp. 337 – 354)

My reaction to this record was that it is, shall we say, slightly implausible to suppose that simply changing from the basic skills TEAMS test to the more challenging TAAS test would lead to statewide changes in teaching in Texas such that within "a year or two" teachers would be focusing not simply on "basic skills" but on "problem solving, higher order thinking skills, making inferences, and drawing conclusions." To test my own reaction against the views of a broader sample of school reform observers, I undertook a "two questions survey of school reform."

So, on Monday, August 16, 1999, I sent a survey via electronic mail to sixteen people, whom I respected as knowledgeable students of school reform initiatives around the country. On August 21, I resent the query to an additional 11 people whose names had been suggested by respondents to my first query. As of September 6, 1999, I had received 10 responses to my questions. Though I do not know what typical response rates are to email surveys of this sort (odd questions posed to busy people in late summer, with no explanation as to their possible import), my own view is that a response rate of 37% (10/27=0.3704) is probably not too bad.

Here is the full text of the email survey, including the two questions posed:

Colleagues: I would like to ask the favor of asking you to answer two questions. Given your professional expertise, I trust the questions will be of some interest. Also, your answers may be of some import. For now, I will not explain the exact reason for my questions, as I would not want it to influence your answers. Imagine a very large school system that has been focusing on basic skills instruction for some years. The focus has been spurred in part by a high stakes test of basic skills. It is assumed that 80-90% of teachers have been covering the basic skills in their instruction.

In light of current educational reform ideas, the system decides that it needs to move beyond basic skills teaching to focus in the future on problem solving, higher order thinking skills, making inferences and drawing conclusions.

In light of this situation, and your expertise in studying school reform, my two questions to you are these:

1. How long would it likely take for this large school system to shift from having 80-90% of teachers teaching basic skills, to having 80-90% of teachers teaching the more advanced skills?
2. What would be the key ingredients required to make such a shift in instruction possible in the time you envision in your answer to the first question?

Please keep your answers brief and email them to me by August 30. In exchange for your kindness in responding to my request, I will compile answers, distribute them to whomever responds, and explain the specific reason that motivates the questions.

The ten scholars who responded to the survey were (in alphabetical order): David K. Cohen, Jane David, Daniel Koretz, Henry Levin, Hayes Mizell, Fred Newmann, Stan Pogrow, Ted Sizer, Adam Stoll, and Anne Wheelock.
Before summarizing what they said in response to the survey, two prefatory points should be added. First, all of these correspondents have generously allowed me to reproduce the full text of their survey responses (see Appendix 9). Second, despite the generosity of these people in responding so quickly (all within three weeks at the end of summer 1999), we did not even attempt to use the survey results in the TAAS trial in September. Inasmuch as lawyers for the State of Texas were already trying to exclude from the trial evidence they had known about for months, Mr. Kauffman advised me that that they might not entirely welcome new evidence from a survey they had not even heard about before the trial began.

As mentioned, all ten responses are reproduced in their entirety in Appendix 9. Here I simply summarize three overall patterns in the ten responses.

*Gentle Chiding.* Half of the respondents (Cohen, Koretz, Pogrow, Stoll and Wheelock) chided me gently for advancing something of a false dichotomy between "basic skills" and advanced or "higher order thinking" skills. I can only plead *mea culpa,* but given the background to the survey explained above, I trust that my oversimplification may be forgiven.

*Shifting the course of large educational systems takes years.* The first question asked "How long would it likely take for this large school system to shift from having 80-90% of teachers teaching basic skills, to having 80-90% of teachers teaching the more advanced skills?" Though all respondents qualified their answers in one way or another, all did provide some sort of time estimate. In brief these were: Cohen, 10 years; David, 10 to 20 years; Koretz, 3 to 4 years; Levin, 2 to 5 years; Mizell, 7 to 8 years; Newmann, At least six years; Pogrow, 2 to 4 years; Sizer, At least 5 years; Stoll, At least 20 years; Wheelock, 10 to 15 years.

Two things strike me about these responses. First is the remarkable variance in responses; from "2 to 4 years" to "at least 20 years" (and even if we throw out these outliers, variance remains nearly as great). This suggests that even among scholars who have studied such matters, we really do not know very much about long it takes to shift the course of large educational enterprises. Second is that the median value seems to fall somewhere in the range of 5 to 10 years. This is of course far longer than the 1 to 2 years presumed by Commissioner Kirby in Texas in 1990.

*Huge resources required.* The second survey question was: "What would be the key ingredients required to make such a shift in instruction possible in the time you envision in your answer to the first question?" Answers to this question were generally far longer than answers to the first question, but in general indicated that a large quantity and range of resources would be needed to change the course of a large educational enterprise, including professional development opportunities for teachers, leadership, community outreach, lower pupil/teacher ratios, more instructional resources, better social services for students, and reform of teacher education institutions. Jane David's summary answer was "massive teacher re-education and powerful recruitment strategies." Henry Levin's answer suggested that significant change in instruction could come about in two to five years, given the following ingredients:

- continuous staff development, continuous support and technical assistance,
- administrative encouragement, intrinsic and extrinsic incentives, public information on results, and a culture of commitment. Add to this transformation of local teacher training programs, careful selection of new teachers, and a strong public relations campaign, and things will move.

Every administrator will have to become a cheerleader.
He then added "The problem is that no district has ever been able to achieve these conditions. Further, this will be competing with basic skills testing that is often high stakes and high visibility promoted by the states."

Adam Stoll wrote, in part:

It's immensely hard to get a critical mass of teachers within a school, let alone a district, to significantly change their practice. I would think getting a majority to exhibit practice that is highly supportive of advanced skill acquisition would be very optimistic, but possibly attainable under optimal circumstances.

I can only imagine having 80-90% of teachers place a lot of emphasis on "teaching the more advanced skills" if some pretty sweeping changes occurred. I think it would take at least 20 years for these changes to begin affecting practice on this scale.

These extracts are really an inadequate summary of the observations offered by survey respondents, so I encourage readers to review their observations, reproduced in full in Appendix 9. Nonetheless, it is clear that very few of the ingredients suggested as needed for large-scale educational reform were provided in Texas in the early 1990s. This suggests why the purported "miracle" of educational reform in Texas is not only largely illusory, but indeed has had widespread negative consequences for both students and educators in the Lone Star state. After recapping the myth of the Texas miracle, I will suggest that this is a lesson from which we should learn. Myopic accountability schemes based on high stakes testing likely will have similarly perverse consequences elsewhere if we do not learn from the unfortunate story of Texas education in the last decade of the 20th century.

8.2 Recapping the Myth

Since the territory covered in this article is extensive, let me try to sum up the journey so far. After an introduction (pointing out among other things that this writer may not be viewed by all as a totally unbiased observer of education in Texas), I summarized the recent history of education and statewide testing in Texas, which led to introduction of the Texas Assessment of Academic Skills (TAAS) in 1990-91. Since then TAAS testing has been the linchpin of educational accountability in Texas, not just for students, but also for educators and schools.

Part 3 recounted how a variety of evidence in the late 1990s led a number of observers to conclude that the state of Texas had made near miraculous educational progress on a number of fronts. Between 1994 and 1998, the percentage of students passing the three grade 10 TAAS tests had grown from 52% to more than 70%. Also, the racial gap in TAAS results seemed to have narrowed. Statistics from the Texas Education Agency showed that over the same interval dropout rates had declined steadily. Finally, in 1997, release of results from the National Assessment of Educational Progress (NAEP) showed Texas 4th graders to have made more progress on NAEP math tests between 1992 and 1996 than those in any other state participating in state NAEP testing. These developments led to a flurry of editorial praise for the apparent educational progress of the Lone Star State. Some went so far as to suggest even that the Texas experience should serve as a model for federal education legislation.

Part 4 began a closer examination of both TAAS and what has been happening in Texas schools over the last several decades. Section 4.1 showed that by any of the
prevailing standards for ascertaining adverse impact, grade 10 TAAS results continue to show discriminatory adverse impact on Black and Hispanic students in Texas. It was also shown that use of TAAS results in isolation to control award of high school diplomas is a clear violation of professional standards concerning appropriate test use. Previously I explained how expert witnesses for the state of Texas had challenged my interpretation or the *Standards for Educational and Psychological Testing*, sponsored by AERA, APA and NCME. In July, 2000, AERA issued a statement that, at least in my view confirms my interpretation of the Standards. (See www.aera.net/about/policy/stakes.htm)

Section 4.2 demonstrated that the passing scores set on TAAS tests were arbitrary, discriminatory and failed to take measurement error into account. Furthermore, analyses comparing TAAS reading, writing and math scores with one another and with relevant high school grades raise doubts about the reliability and validity of TAAS scores. Finally, it was demonstrated how a sliding scale approach (taking into account both test scores and grades) could be applied in a more professionally sound and less discriminatory manner.

Stepping back from the arcane technology of standardized testing, Part 5 discussed problems of missing students and other mirages in Texas. First, patterns of student enrollment in Texas between 1975 and 1999 were examined by studying rates of progress from grade 9 to high school graduation, grade to grade progression ratios, and grade 6 to high school graduation rates. Without trying to summarize results of all of those analyses here, let me mention just some of the substantive findings from these analyses. In 1990-91, Black and Hispanic high school graduates relative to the number of Black and Hispanic students enrolled in grade 9 three years earlier fell to less than 0.50 and this ratio remained just about at or below this level from 1992 to 1999 (the corresponding ratio had been about 0.60 in the late 1970s and early 1980s). This finding indicated that only 50% of minority students in Texas have been progressing from grade 9 to high school graduation since the initiation of the TAAS testing program.

Subsequent analyses of progression ratios for all the grades indicated that the rates of Texas students being denied promotion from grade 9 to 10 have changed sharply over the last two decades. From 1977 until about 1981 rates of grade 9 retention were similar for Black, Hispanic and White students, but since about 1982, the rates at which Black and Hispanic students are denied promotion and required to repeat grade 9 have climbed steadily, such that by the late 1990s, nearly 30% of Black and Hispanic students were "failing" grade 9 and required to repeat that grade.

This finding led to a third series of analyses examining rates of progress from grade 6 and grade 8 to high school graduation. It was found that the rate of progress from grade 6 to high school graduation fell from about 0.75 in 1990 to less than 0.70 for White students and from about 0.65 to 0.55 for minority students. (The rate for minority students started to climb above 0.60 only in 1997, the year in which Texas was forced to raise the passing score on the GED high school equivalency tests).

Since all this discussion of rates and ratios may well obscure what is happening – or not happening – to large numbers of children in Texas, let us take one last look at the grade enrollment data for Texas. This time I show simply numbers of students, not ratios or percentages. Figure 8.1 shows progress from grade 6 to high school graduation 6.5 years later for the Texas high school classes of 1982 to 1999 simply in terms of numbers of students (that is, total numbers of Black, Hispanic and White students).
Also shown in this figure is the difference, that is the numbers of students who do not make it from grade 6 to high school graduation 6.5 years later. As can be seen, the numbers of children lost between grade 6 and high school graduation in Texas were in the range of 50 to 60 thousand for the classes of 1982 to 1986. The numbers of lost children started to increase for the class of 1987 and jumped too almost 90 thousand for the class of 1991. For the classes of 1992 through 1999, in the range of 75 to 80 thousand children are being lost in each cohort. (For readers who may have not waded through all of the previous parts of this very long article and simply skipped to this conclusion, it is worth noting that as discussed in Part 7, these estimates are probably conservative, since there has been a net in-migration of people into Texas in the last two decades.

Cumulatively for the classes of 1992 through 1999, there were about 2.2 million enrolled in grade 6 (in the academic years 1984-85 through 1992-93). The total number graduating from these classes was about 1.5 million. In other words, for the graduating classes of 1992 through 1999, around 700,000 children in Texas were lost or left behind before graduation from high school.

Section 5.4 of the article examined cumulative rates of grade retention in Texas. These are almost twice as high for Black and Hispanic students as for White students. The next section (Section 5.5) reports on estimates of dropouts by grade. It was found that most dropouts occur between grade 9 and 10 (about 16% of Black and Hispanic students and 8% of White students) but that another 6 to 10 percent dropout after grade 10 and also after grade 11. This portion of the article also shows the way in which apparent increases in grade 10 TAAS pass rates tend to disappear, if they are based not on numbers of students taking TAAS in the spring of grade 10, but instead on fall grade 9 or even fall grade 10 enrollments.

Having been alerted to the fact that some portion of the gains in grade 10 TAAS pass rates were illusory, in Section 5.6 I next sought to estimate the numbers of students taking the grade 10 tests who were classified as "in special education" and hence not counted in schools' accountability ratings. As reported in Section 5.6, the numbers of
such students nearly doubled between 1994 and 1998.

In the closing portion of Part 5, I sought to estimate what portion of apparent gains in TAAS pass rates might be due to such forms of exclusion. It was estimated that a substantial portion, but probably less than half of the apparent increases in TAAS pass rates in the 1990's are due to such exclusions.

In Part 6 of this article, I sought to summarize the views of educators in Texas about TAAS, based on three statewide surveys of educators. These surveys were undertaken entirely independently, and surveyed somewhat different populations of educators. General findings from this review were as follows:

1. Texas schools are devoting a huge amount of time and energy preparing students specifically for TAAS.
2. Emphasis on TAAS is hurting more than helping teaching and learning in Texas schools.
3. Emphasis on TAAS is particularly harmful to at-risk students.
4. Emphasis on TAAS contributes to retention in grade and dropping out of school.

Survey results indicated that the emphasis on TAAS is contributing to dropouts from Texas schools not just of students, but also teachers. In one survey, reading specialists were asked whether they agreed with the following statement:

It has also been suggested that the emphasis on TAAS is forcing some of the best teachers to leave teaching because of the restraints the tests place on decision making and the pressures placed on them and their students.

A total of 85% of respondents agreed with this statement. In another survey, teachers volunteered comments such as the following: "Mandated state TAAS Testing is driving out the best teachers who refuse to resort to teaching to a low level test!"

The penultimate portion of this article, Part 7, reviews a variety of additional evidence about education in Texas. Five different sources of evidence about rates of high school completion are compared and contrasted. In an effort to reconcile sharp differences apparent in these sources, a review of statistics on numbers of students, in Texas and nationally, taking the Tests of General Educational Development (GED) was undertaken. People take the GED tests in order, by achieving passing scores, to be awarded high school equivalency degrees. The review of GED statistics indicated that there was a sharp upturn in numbers of young people taking the GED tests in Texas in the mid-1990s.

This finding helps to explain why the TEA statistics on dropouts are misleading. According to TEA accounting procedures, if students leave regular high school programs to go into state-approved GED preparation programs, they are not counted as dropouts. As Greene (1998) observed:

[A]n important misleading feature of the [TEA] reported drop-out rates is that they exclude students who were transferred to approved alternate programs, including drop-out recovery programs. If the students in these drop-out or other alternative programs subsequently drop out, it is not counted against the district. This is like reporting death rates at hospitals where you exclude patients transferred to intensive care units.

If we put aside the TEA-reported dropout rates as misleading, differences in other sources of evidence on rates of high school completion in Texas appear reconcilable.
NCES reports (based on CPS surveys) indicate that the rate of high school completion among young people in Texas in the 1990s was about 80%. This would imply a non-completion (or dropout) rate of 20%. Initially this would appear markedly lower than the non-graduation rate of at least 30% derived from my analyses of TEA data on enrollments and graduates. But the CPS surveys count as high school completers, those who receive a regular high school diploma and those who receive a GED high school equivalency degree. So it seems clear that a convergence of evidence indicates that during the 1990s, slightly less than 70% of students in Texas actually graduated from high school (e.g., 1.5 million/2.2 million = 0.68). This implies that about 1 in 3 students in Texas in the 1990s dropped out of school and did not graduate from high school. (Some of these dropouts may have received GED equivalency degrees, but as discussed in Part 7, GED certification is by no means equivalent to regular high school graduation).

Section 7.2 examined patterns of retention in grade 9 and high school completion among states for which such data are available. Results indicated that there is a strong association between high rates of grade 9 retention and low rates of high school completion (specifically, results suggested that for every 10 students retained to repeat grade 9, about seven will not complete high school).

Part 7.3 examined SAT scores for Texas students as compared with national results. Evidence indicates that at least as measured by performance on the SAT, the academic learning of secondary school students in Texas has not improved since the early 1990s, at least as compared with SAT-takers nationally. Indeed results from 1993 to 1999 on the SAT-M indicate that the learning of Texas student has deteriorated relative to students nationally (and this result holds even after controlling for percentage of high school graduates taking the SAT).

Part 7.4 revisited NAEP results for Texas. Results for eight state NAEP assessments conducted between 1990 and 1998 were reviewed. Because of the doubtful meaningfulness of the NAEP achievement levels, NAEP results for Texas and the nation were compared in terms of NAEP test scores. In order to compare NAEP results with those from TAAS, the "effect size" metric (from the meta-analysis literature) was employed. This review of NAEP results from the 1990s, showed that grade 4 and grade 8 students in Texas performed much like students nationally. On some NAEP assessments Texas students scored above the national average and on some below. In the two subject areas in which state NAEP assessments were conducted more than once during the 1990s, there is evidence of modest progress by students in Texas, but it is much like the progress evident for students nationally. Reviewing NAEP results for Texas by ethnic group, we see a more mixed picture. In many comparisons, Black and Hispanic students show about the same gain in NAEP scores as White students, but the 1998 NAEP reading results indicate that while White grade 4 reading scores in Texas have improved since 1992, those of Black and Hispanic students have not. More generally, however, the magnitudes of the gains apparent on NAEP for Texas fail to confirm the dramatic gains apparent on TAAS. Gains on NAEP in Texas are consistently far less than half the size (in standard deviation units) of Texas gains on state NAEP assessments. These results indicate that the dramatic gains on TAAS during the 1990s are more illusory than real. The Texas "miracle" is more hat than cattle.

The final portion of the penultimate portion of this article (Section 7.5) provided a brief review of other evidence concerning the state of education in Texas. Perhaps the most striking portion of this review were results from the Texas Academic Skills Program or TASP test during the 1990s. Between 1994 and 1997, TAAS results showed a 20% increase in the percentage of students passing all three exit level TAAS tests (reading, writing and math). But during the same interval, TASP results showed a sharp
decrease (from 65.2% to 43.3%) in the percentage of students passing all three parts (reading, math, and writing) of the TASP college readiness test.

8.3 Testing and Accountability

What might be the broader lessons from the Texas myth for education elsewhere? Surely there are many different ones that might be read into this story (such as the need to be wary of the party line emanating from large bureaucracies, which education in Texas seems to have become; and the importance of comparing alternative forms of evidence in order to begin to get at the truth about large and complex enterprises). But in closing, I comment briefly on only three of what I view as the broader lessons from the Texas myth story.

Aims of Education. The Texas myth story surely helps remind us of the broader aims of education in our society. The dramatic gains apparent on TAAS in the 1990s are simply not born out by results of other testing programs (such as the SAT, NAEP and TASP). But quite apart from test scores, surely one of the main outcomes of pre-collegiate education is how many students finish and graduate from high school. By this measure of success, surely the Texas system of education in which only two out of three young people in the 1990s actually graduated from high school should not be deemed a success, much less a miracle.

Testing and Accountability. The TAAS testing program in Texas seems to have been spawned mainly by a yen for holding schools "accountable" for student learning. It is an unfortunately common manifestation of what has come to be called in the last several decades "outcomes accountability." As suggested above, however, quite apart from test scores, surely one of the most important outcomes of public education is how many young people finish schooling and graduate from high school. And this reminds us of the broader meaning of the term accountability (Haney & Raczek, 1994). In its broader meaning the word accountability refers to providing an account or explanation not just of consequences, but of conduct. The Texas myth story, it seems to me, reminds us of how vital it is when judging educational endeavors to return to the root meaning of the word accountability and inquire into conduct as well as consequences.

It is of course always possible to come up with some sort of bureaucratic scheme, as in Texas, for weighing various sorts of data about schools and coming up with some kind of summary judgment about their quality. But anyone who believes in the rationality of such approaches has forgotten the old paradox of value from the field of economics. The paradox refers to the fact that many obviously useful commodities, such as air and water, have very low if any exchange values, whereas much less useful ones such as diamonds and gold, have extremely high value. According to Schumpeter’s (1954) *History of economic analysis*, it was recognized as early as the 16th century, by "scholastic doctors" and natural philosophers that the exchange value or price of commodities derived not from any inherent characteristics of the commodities themselves but from their utility or "desiredness" and relative scarcity. Without wandering into a digression on the field of economic theory (concerning which I am an absolute amateur anyway), let me simply mention how this paradox was resolved by Kenneth Arrow. In 1950, Arrow published what has come to be known as his "impossibility theorem," in an article modestly titled "A difficulty in the concept of social welfare." In this article, Arrow proved mathematically that if there are at least three alternatives which members of society are free to order in any way, any social welfare function yielding an ordering based on those preferences violates one of three rational conditions (as long as trivial and dictatorial methods of aggregation are excluded). In short Arrow’s "impossibility
Hazard of High Stakes Testing. More than anything though, the Texas miracle story shows us the hazards of high stakes testing. It is, of course, possible to impose a "whips and chains" test-based accountability system on schools (as Schrag, 2000, described the Texas approach). Yet the Texas miracle story shows us the need to return standardized testing to its rightful place, as a source of potentially useful information to inform human judgment, and not as a cudgel for implementing education policy.
Notes

A previous version of this paper was presented at the Annual Meeting of the American Educational Research Association, New Orleans, April, 2000.

1. The normal legal citation for Judge Prado's decision in the TAAS case is GI Forum Image De Tejas v. Texas Education Agency, 87 F. Supp. 667 (W.D.Tex. 2000). However, since this citation only recently became available, in the body of this paper I cite Judge Prado's decision as Prado, 2000.

2. The volume 2, number 2 issue of *The Scholar: St. Mary's Law Review on Minority Issues* has recently published major portions of the reports of eight experts who testified in the TAAS trial, including portions of both my original (Haney, 1998) and supplementary (Haney, 1999) reports concerning the case.

3. I do not know how many schools have been taken over by the state, but I am aware that the TEA took over control of the Wilmer-Hutchins district in 1996 because of poor performance (Wertheimer, 1999).

4. At least one independent analyst has found that the equating of TAAS forms has not been successful. In a study commissioned by the Tax Research Association of Houston and Harris County (TRA), Sandra Stotsky analyzed TAAS reading tests for 1995 through 1998 and found that the grade 4, 8 and 10 TAAS reading tests for these years and grades were not comparable in difficulty (see, Stotsky, 1998).

5. In the second report for the TAAS case (Haney 1999), I also applied the 80% rule to results for three different grade 10 TAAS tests (reading writing and math). Writing test results for Blacks and Hispanics have generally not fallen below 80% of the White pass rates, but TAAS math test results consistently have.

6. The latest version of the joint test Standards was issued in 1999, after the TAAS case and my work on it, were under way. Therefore, here I cite both 1985 and 1999 versions of the Standards. Where pertinent, I also document how specific provisions changed between 1985 and 1999.

7. The corresponding 1985 standard read: Standard 8.12 In elementary or secondary education, a decision or characterization that will have a major impact on a test taker should not automatically be made on the basis of a single test score. Other relevant information for the decision should also be taken into account by the professionals responsible for making the decision. (APA, AERA & NCME, 1985, p. 54)

8. MALDEF attorneys sought to have the Heubert & Hauser report entered as evidence in the TAAS trial, but after attorneys for the state of Texas objected, the judge refused to allow the NRC report entered as evidence in the case. In a symposium on the GI Forum case at the Annual Conference of the Council of Chief State School Officers, Snowbird Utah, June 17, 2000, I asked Geoffrey T. Amsel, the lead lawyer for the State of Texas in the case, why in the world he had sought to have NRC report excluded from evidence in the case. His public response? "I was just trying to be a pain in the ass."

9. The corresponding passages from the 1985 Standards are: Standard 6.9 When a specific cut score is used to select, classify, or certify test takers, the method and
rationale for setting that cut score, including any technical analyses, should be presented in a manual or report. When cut scores are based primarily on professional judgment, the qualifications of the judges also should be documented. (AERA, APA & NCME, 1985, p. 34) And 1985 Standard 2.10 specifies that "standard errors of measurement should be reported for score levels at or near the cut score" (p. 22).

10. It is worth mentioning that since 1990 considerable literature has been published on methods for setting passing scores on tests (for example, Gregory Cizek, Setting passing scores, *Educational Measurement: Issues and Practice*, Summer 1996, pp. 20-31). However in discussing the setting of passing scores on TAAS in 1990, it seems reasonable to focus on literature that was prominently available before that year.

11. Part 6.2 below provides more explanation on how this survey was undertaken. Haney, *Myth of the Texas Miracle*, v. 4, July 28, 2000, p. 58.

12. I was able to assemble this data set thanks to the generous assistance of Dr. Ed Rincon of Rincon Associates and Terry Hitchcock of the Texas Education Agency (TEA).

13. When a graph like Figure 5.2 was presented during the TAAS trial (the same, except that it did not include 1998-99 data), the pattern was sufficiently startling that Judge Prado interjected exactly this question, "What happened?"

14. The original Table 6.1 in the NRC report contained several printing errors, but a corrected version has been released.

15. It should be explained that the TEA data cited shows slightly different numbers of students taking the three portions of the grade 10 TAAS (reading, writing, and math) in any given administration. To derive the results shown in Table 5.5, I calculated the number of special education students taking each portion of the TAAS in each year and then averaged the numbers and percentages taking each portion.

16. We should acknowledge that this response rate of less than 15% was certainly less than ideal. One likely reason for the low response rate is that we were able to mail the survey only one week before the last week of the 1998-99 Texas school year. One respondent even spontaneously chided us for sending a survey that arrived during such a hectic time in the school year. Because of this timing we were unable to send follow-up letters to non-respondents.

17. ARD stands for Admission, Review and Dismissal, the name of the Committee in Texas schools that oversees special education designations and plans.

18. Recently, thanks to a suggestion of Jeff Rodamar, I have become aware of a set of short papers on the web site of the Texas Public Policy Foundation (http://www.tppf.org/). One paper, by John Pisciotta summarizes two 1996 surveys of satisfaction and dissatisfaction of teachers in Texas. The report states: "Overall, this report indicates that with all the Texas public education reforms of recent years, the environment for Texas professional educators has not improved. One key finding is that public school teachers did not generally believe the teaching quality in their schools was improving. When asked if quality of teaching at their school had improved compared to five years ago, only 39% of public school teachers said yes. Almost a third believed teaching quality was worse than five years ago. In contrast, 71% of private school teachers saw their schools as better than five years ago. Social promotion, passing students from one grade to the next without adequate academic achievement, was another topic of the surveys. Public school teachers viewed social promotion as a widespread problem. Over half of the public school teachers indicated that social promotion was a problem at their school,
compared to 29% of private school teachers. The most central question relating to teacher attrition was: Are you seriously considering leaving the teaching profession? For public school teachers, 44% said they were. Only 28% of the private-school teachers were seriously considering leaving the profession. As the major reason for leaving, private school teachers cited inadequate financial compensation. Public school teachers cited poor working conditions as their major reason for leaving." (http://www.tppf.org/, accessed 5/7/00).

19. In trying to track down possible sources of discrepancies in Texas dropout rates, I talked with Phil Kaufman of MPR of Berkeley California. Among other things he explained that the CPS data gathering began to use computer assisted telephone interviewing in 1994, and hence it is hazardous to compare CPS results from before and after that date.

20. In order to further explore this issue, I consulted with a number of scholars who have previously analyzed CPS, data including Robert Hauser, Phil Kaufman, Richard Murnane, Duncan Chaplin and John Tyler. What I conclude from these consultations is that for a variety of reasons, one needs to be wary of dropout rate estimates based on CPS data. See, for example, Hauser, 1997; Chaplin, 1999.

21. The very next sentence after the passage quoted here says "Consequently, GED graduates in 1997 and beyond must meet or surpass the performance of the top two-thirds of traditional graduating high school seniors." Obviously this statement is mistaken. What was meant was that the new GED passing standard raised the minimum scores such that instead of exceeding the performance of 25% of the norm group of high school seniors, the new minimum was equal to or surpassed the performance of 33% of the norm group.

22. A minor mystery appeared when it was learned that 15 to 20% of GED takers in Texas were only 16 or 17 years of age. GED annual reports indicate that the minimum age for taking the GED in Texas is 18. So I called the Office of Continuing Education in the Texas Education Agency (512-463-9292, 6/1/00). It was readily explained that people can take the GED in Texas below age 18 if they have a letter from a parent, parole officer, or judge. In a personal communication (6/8/00), John Tyler generously told me how to solve another mystery. GED statistics from the TEA are slightly different than those reported by GEDTS, apparently because TEA tends to report GED statistics in terms of GED certificates actually awarded, whereas GEDTS also reported numbers who pass the GED tests.

23. The only jurisdiction with a larger drop in its passing rate in 1997 was American Samoa, where only 30 people were tested in 1997.

24. To be clear, the new GED passing standard in Texas was more difficult than the pre-1997 Texas standard. It appears to be much lower than the passing standard on TAAS. Though I have been unable to locate any studies comparing the difficulty of the TAAS and GED tests, according to Barasch et al. (1990, p. 9) "To be successful in passing the GED in most states, a candidate must get a total minimum standard score of 225 on the five tests, with no score less than 35 on any single test. In general this means that a candidate who answers just over half of the questions in each test will get a passing score." As we have seen, in Texas until 1997, people could pass the GED with a total standard score of only 200.

25. For more recent evidence on economic returns to earning the GED, see Murnane, Willett, & Tyler, 2000.

26. It is worth noting that analyses of grade enrollment data in part 5.5 above suggest that Murdock et al.'s estimate of the 1-2% annual in-migration rate for the Texas population appears to hold for the school age population in the 1990s. For example,
referring to Table 5.3, if we average the % difference between predicted and actual grade enrollments for 1996-97 for grades 2-7 where retention in grade is quite uncommon, we get a little over 1% across the three ethnic groups. Note too that to the extent that Hispanic in-migration is greater than White in-migration, as Murdock et al., indicate, so too will the Hispanic-White gap in dropout rates be underestimated.

27. The correlation between these two variables is -0.51, statistically significant at the 0.05 level, even with the small sample of states for which grade 9 retention rates are available. Also, I suspect that for Arizona, the outlier data point in Figure 7.3, data on high school completion may be unreliable. If we replace the 77.1 % high school completion rate for 1996-98 for Arizona with the rate of 83.8% that Kaufman et al. (1999) report for Arizona for 1993-95, the correlation changes to -0.7. And if we simply delete the Arizona case, the correlation is -0.80. If the Arizona case is deleted, the regression of HS completion rate on grade 9 retention rate is HSC = 95.6 - 0.69G9R (R2 = 0.657) This suggests that for every 10 students retained to repeat grade 9, about seven will not complete high school. Given this regression equation, the predicted rate of high school completion for Texas would be 83.3, but the actual rate is about three points lower, at 80.2.

28. This finding is particularly significant given that previous research has shown that quantitative test scores are more sensitive to school experiences than are verbal test scores (Haney, Madaus & Lyons, 1993).

29. The exception to this pattern is that since it was decided that a student must achieve at least a 2 score on the written composition in order to pass the TAAS writing test, a composition score of "1" plus 27 or more multiple-choice items is truncated to a scale score of 1499, which is one point below mastery.

30. Rodamar (2000) also has an interesting summary of how college-going in Texas has changed between 1994 and 1996 (see exhibit 15, p. 21).

References


Association, Chicago, IL:


Texas House of Representatives House Research Organization (1999). *The Dropout Data*


The Myth of the Texas Miracle in Education: Appendices

- Appendix 1: Testing & Teaching Survey Form
- Appendix 2: Testing & Teaching Survey Comments
- Appendix 3: Judge Prado Decision in GI Forum Case
- Appendix 4: Plaintiffs' Post Trial Brief
- Appendix 5: Defendants Response to Plaintiff's Post Trial Brief
- Appendix 6: Summary Comment on Judge Prado's Decision in the GI Forum Case
- Appendix 7: Texas Enrollments by Grade and Race (1975-1999)
- Appendix 8: Minutes of the Texas State Board of Education Meeting, July 1990
- Appendix 9: Responses to Two Questions Survey