

**Legal Challenges to High-Stakes Testing:  
A Case of Disparate Impact in Michigan?**

Paper Presented at the Annual Meeting of the  
American Educational Research Association  
Seattle, Washington  
April, 2001

Donald E. Heller  
Assistant Professor of Education  
Center for the Study of Higher and Postsecondary Education  
University of Michigan School of Education  
610 East University, 2108D SEB  
Ann Arbor, MI 48109-1259  
734-647-1984  
Fax: 734-764-2510  
dheller@umich.edu

Douglas T. Shapiro  
Doctoral Candidate  
Center for the Study of Higher and Postsecondary Education  
University of Michigan School of Education  
610 East University, 2117 SEB  
Ann Arbor, MI 48109-1259  
734-764-9472  
dtshapir@umich.edu

The authors wish to acknowledge the research assistance of  
Christopher Rasmussen in the preparation of this study.

© 2001, Donald E. Heller and Douglas T. Shapiro

# **Legal Challenges to High-Stakes Testing: A Case of Disparate Impact in Michigan?**

## **Background on State Scholarships and the Michigan Merit Award Scholarship Program**

Since the passage of the Higher Education Act of 1965, which implemented the Educational Opportunity Grant program (the precursor to Pell Grants), financial need has been the primary criterion used in awarding publicly-funded scholarships for college. The states followed the lead of the federal government in awarding grants to needy students. State-funded scholarships grew with the creation of the State Student Incentive Grant (SSIG) program in 1972, which provided matching federal funds to states that created their own scholarship programs. The great majority of these state scholarship funds were awarded based on financial need.

Since the 1980s, however, the use of financial need as the basis for awarding scholarships by the states has been eroding. Between 1982 and 1999 spending on need-based scholarships for undergraduates by the states increased 7.3 percent annually, while spending on merit programs increased at a 12.7 percent annual rate. The proportion of state grants awarded based on merit has risen from 9.0 percent to 18.6 percent during this period (Heller, in press).

There are numerous examples of state merit scholarship programs that have been created in recent years. The most well known is the Helping Outstanding Pupils

Educationally (HOPE) program in Georgia. Begun in 1993, it is now the largest state-run merit scholarship program in the country, awarding \$189 million in the 1998/1999 academic year (Heller, in press). The criterion used for the awarding of HOPE scholarships is the attainment of a B average (3.0 on a 4.0 scale) in a selection of high school core curriculum subjects (Mumper, 1999).

Looking to duplicate the popularity of the HOPE program among Georgia voters,<sup>1</sup> Governor John Engler of Michigan announced in his 1999 State of the State Address that he would introduce legislation to create the Michigan Merit Award, “for all Michigan high school graduates who master reading, writing, math and science” (Engler, 1999). Engler proposed that the program be funded from a portion of the state’s share of tobacco settlement funds. The bill quickly passed through the legislature and the Michigan Merit Award Scholarship Act was signed into law on June 30, 1999. The stated goal of the Act was “to increase access to postsecondary education and reward Michigan high school graduates who have demonstrated academic achievement” (“Michigan merit award scholarship act,” 1999, p. 2).

The Act provides scholarships of up to \$2,500 for students who score at Level 1 (exceeds Michigan standards) or Level 2 (meets Michigan standards) on all four portions of the Michigan Educational Assessment Program High School Tests (MEAP HST). The MEAP tests are a criterion-referenced test designed to measure knowledge of the state’s designated curricular frameworks. The tests are given in four subject areas: mathematics, reading, science, and writing. Although the vast majority of scholarship recipients qualify through the MEAP test, the legislation also provides an

alternative path for qualifying for the scholarships. To qualify under this alternative path, students must: 1) take all four subject area tests; 2) receive a score of Level 1 or 2 on at least two of the tests, and 3) score in the top 25 percent nationally on the SAT 1, ACT, or ACT WorkKeys tests.<sup>2</sup> All students in Michigan, regardless of family income or other characteristics, are eligible for the awards.<sup>3</sup> Students from the graduating class of 2000 were the first cohort eligible for the scholarships. Funding for the program was set by the legislation at 30 percent of the tobacco settlement total in fiscal year 2000, 50 percent in 2001, and 75 percent in 2002 and subsequent years.

In June 2000, a coalition of groups headed by the American Civil Liberties Union of Michigan filed suit against the state in federal court, asking for a preliminary injunction to block the implementation of the Michigan program (*White et al. v. Engler et al.*).<sup>4</sup> The complaint alleges that the Michigan scholarship program “discriminates against high school students on the basis of race, ethnicity, and educational disadvantage” and is thus in violation of Title VI of the Civil Rights Act of 1964 and the Equal Protection Clause of the Fourteenth Amendment (p. 1). As noted in the complaint, “Title VI ... in clear and unmistakable terms prohibits agencies from implementing policies which have a disparate impact on the basis of race or ethnicity” (p. 14). In legal terms, disparate impact refers to a lesser standard of discrimination that requires only a showing of discriminatory effect in the implementation of a policy, not actual discriminatory intent in the creation of the policy.

The ACLU’s complaint in Michigan does not challenge the offering of the scholarship money itself, but rather the state’s method of determining qualification for

the scholarships — the use of the MEAP test as a “high-stakes” single criterion measure. Students from minority and lower-income groups nationwide have long been known to underperform on standardized tests (Hedges & Nowell, 1999; Orfield & Wald, 2000; and Urdan & Davis, 1998). But the use of score cutoffs on such tests to make educational decisions like awarding college scholarships translates that underperformance into the underrepresentation of minority groups among scholarship qualifiers. It is this alleged disparate impact in the awarding of the Michigan Merit scholarships that forms the foundation of the ACLU’s complaint against the state.

This study analyzes data from the first cohort of students eligible for the Michigan merit scholarships and addresses these questions:

1. What is the relevant case law and theoretical frameworks related to the impact of testing in educational settings?
2. What is the racial and ethnic distribution of the first cohort of students eligible for the scholarships, and is there *prima facie* evidence of disparate impact?
3. Is the scholarship program likely to achieve the legislated goal of increasing access to higher education in the state?

### **Legal and Theoretical Frameworks**

Disparate impact in the context of educational testing refers to a disproportionate disadvantage imposed on a group of students from the application of a facially neutral criterion or test, where the criterion is not justified by educational necessity (Kimmel,

2001). If the disadvantaged group is defined by race, sex, national origin or disability, the test may be subject to disparate impact claims under the Equal Protection Clause of the Fourteenth Amendments to the U.S. Constitution, Title VI of the Civil Rights Act of 1964, Title IX of the Education Amendments of 1972, Title II of the Americans with Disabilities Act of 1990, or §504 of the Rehabilitation Act of 1973. Under these statutes' implementing regulations (although not under the statutes themselves), it has generally been established that there is no requirement that a plaintiff show discriminatory intent (Vaseleck, 1994). Instead, applications of educational tests that result in disparate impacts have been held to be inappropriate under several possible circumstances, including: when the test is used for a purpose for which it was not designed or validated, when the test score is the sole criterion for an educational decision, when there is no educational basis for establishing a cutoff score, and when the test predicts or performs differently for members of different groups (Kimmel, 2001).

The key legal precedent for such claims is (*Sharif v. New York State Education Department*, 1989). In this case, a group of female high school students challenged the State's use of a standardized test for awarding Regents Scholarships and Empire State Scholarships of Excellence. The challenge was based on the test's disparate impact against female students, in violation of the equal protection clause and Title IX. The scholarships had been established by the state legislature to recognize and award superior high school achievement, providing an award of \$1,000 to top students. The State Education Department had, over the years, experimented with a number of different criteria for awarding the scholarships. These included a Regents scholarship

examination specially designed to measure achievement in college preparatory courses, SAT scores alone and a combination of SAT scores and high school GPAs. In 1989, largely due to budgetary limitations, the State decided to rely solely on the SAT, a move that resulted in male students receiving substantially more scholarships than females, both in absolute terms and relative to the proportions of test-takers. The plaintiffs then brought the nation's first disparate impact challenge to an educational testing practice under Title IX, claiming that the use of the SAT as a sole criterion for awarding the scholarships violated the statute. The federal district court agreed, enjoining the state's use of SAT scores, for two main reasons: first, because the test had been neither designed nor validated for measuring high school achievement, the construct that the legislature had intended to award; and second, because the statistical evidence presented at trial showed that the SAT consistently underpredicts first-year college performance for women, resulting in disparate impact on the basis of sex.

The court in *Sharif* also maintained that the three-step process for proving disparate impact cases that had been established under Title VII of the Civil Rights Act of 1964 (beginning with *Griggs v. Duke Power Co.*, 1971), was also applicable to educational testing cases under Title IX. First, plaintiffs must show that a facially neutral practice has a disproportionate effect. Courts have generally applied two tests to assess whether a disparate impact exists. The first is whether a "statistical analysis shows that the success rate for members of a protected class is significantly lower (or the failure rate is significantly higher) than what would be expected from a random distribution" (Heubert & Hauser, 1999, p. 59). The second test is "if the success rate of a

protected group is less than four-fifths, or 80 percent, of the rate at which the most highly selected group (usually whites or males) is selected" (p. 59, n15).

The use of statistical evidence was constricted, however, in *Wards Cove Packing Co. v. Antonio* (1989), a case of racial discrimination in the workplace under Title VII, in which the U.S. Supreme Court rejected the possibility that statistics revealing a racially imbalanced result might alone be probative of disparate impact. Instead, the Court required a more narrow statistical comparison between the racial composition of the result and the composition of the relevant base group, in this case the pool of qualified applicants. It has been suggested, for example, that the statistical evidence produced in *Sharif* would nonetheless have met the *Wards Cove* standard (had it been in place at that time) since the evidence established not only that female students were underrepresented among scholarship recipients relative to their representation among test-takers, but also relative to their representation among top high school achievers when that achievement was measured by high school GPAs ("Civil rights – Disparate-impact doctrine," 1990).

In the second step, the burden shifts to defendants to show a substantial legitimate justification for their testing practices. This burden has been interpreted as a requirement to demonstrate the "educational necessity" of the practices (*Groves v. Alabama State Bd. of Educ.*, 1991), in particular, that the test has been designed and validated for the purpose intended, and that the specific cutoff scores employed (if any) bear a significant relationship to the educational competencies intended. Two examples serve to illustrate this standard. In *Sharif*, plaintiffs successfully argued that although

the explicit legislative intent of the scholarships was to reward high school achievement, the SAT had been designed not to measure achievement, but to predict college success, and that it had never been validated for the purpose of measuring achievement. In *Groves*, the use of a cutoff score on the ACT as a requirement for admission to a teacher training program was struck down because the state had failed to show that the cutoff was an educationally justified measure of the minimum competency necessary for becoming a teacher (Kimmel, 2001).

Finally, in the third step, the plaintiff may still prevail by showing that an equally effective alternative method exists and has a less discriminatory impact. Since New York State had experimented with several different methods of determining eligibility for Regents Scholarships in the past, the plaintiffs in *Sharif* were able to provide direct statistical evidence that the use of a combination of SAT scores and high school GPAs would have a less discriminatory effect.

Although *Sharif* set the standard for disparate impact cases in educational testing overall, specific claims of racial discrimination in educational testing using disparate impact theory under Title VI of the 1964 Civil Rights Act have another precedent in *Cureton v. NCAA* (1999). Plaintiffs in *Cureton* claimed that the initial eligibility rule of the NCAA, which required students to attain a fixed cutoff score on either of two standardized tests (the SAT or the ACT) as a condition of eligibility to participate in intercollegiate athletics, has an unjustified disparate impact against African-Americans. A federal district court agreed, declaring illegal the minimum standardized test score requirement, largely on the basis of plaintiff's third step demonstration of three

alternative practices that would result in less racial disproportionality while still serving the NCAA's goals (for example, a combined scale of test scores and high school grades, with no fixed cutoffs for either measure independently, see Kimmel, 2001).

This study focuses primarily on developing the statistical evidence required in the first step of the process defined in *Sharif*. A significant body of past educational research has been conducted in this area, establishing the general relationship between students' socioeconomic characteristics and the results of standardized tests. In particular, studies of high school students that have looked at student characteristics such as race, ethnicity, or socioeconomic status (SES) have consistently found strong relationships between those constructs and performance on standardized tests. Significant achievement gaps between White and Asian American students on one hand, and Hispanic and African American students on the other, or between high SES students and low SES students, have been identified. These gaps persist regardless of what specific learning outcome is measured, or whether the analysis is conducted at the level of individual students within schools or at the level of the schools themselves.

At the national level, the standard reports come from the National Assessment of Educational Progress (NAEP), which has been administered by the U.S. Department of Education for students in grades 4, 8 and 12 since 1969. Comparing student racial subgroups on the reading component of the 1998 NAEP, for example, Donahue, Voekl, Campbell, and Mazzeo (1999) reported that the average reading scores for White students was higher than that for African American, Hispanic, and Native American students at all grade levels tested. Similar results have been found for the other

components of the test, including mathematics, science, and writing (Greenwald, Persky, Campbell, & Mazzeo, 1999; O'Sullivan, Reese, & Mazzeo, 1997; Reese, Miller, Mazzeo, & Dossey, 1997). A number of researchers have synthesized data from various Education Department reports to draw further conclusions. Lee (1998) for example, organized achievement gaps from the NAEP mathematics test into four dimensions, pairing the within- and between-school levels with each of the categories of race and SES. His results, focusing on state policy correlates of gaps from the 1992 NAEP, indicate that while the existence of the gaps is universal, their size and the significance of the independent variables varies among different states. Sedlacek (1995) used hierarchical linear modeling to simultaneously weigh both within-and between-school effects, finding similar racial linkages to achievement across subject areas, and when controlling for gender, SES, and course-taking patterns. Bruschi and Anderson (1994) looking specifically at science achievement on the 1990 NAEP, also confirmed large differences between White, Hispanic, and African American groups across science content areas.

Our research extends these results into the specific arena of the performance of different racial groups of Michigan high school students who attempted to qualify for a Michigan Merit Scholarship.

## Methodology

The data used in this study are from the first cohort of students eligible for the Michigan merit award scholarships. The data were provided by the Michigan Merit Award Board. Additional data were acquired from the Michigan School Report (MSR), which includes data on enrollments, graduates, and college participation rates of public high schools in the state (Michigan Department of Education, 1999).

The second research question – concerning the racial and ethnic distribution of the scholarship awards – is answered using bivariate analysis techniques. We compare the distribution of students who qualified for the scholarship from each demographic category with the distribution of students who attempted to qualify. The distribution of students with different background characteristics among scholarship qualifiers indicates which students in the state are benefiting most from the scholarship program. Over- or under-representation of scholarship qualifiers relative to those attempting to qualify indicates an imbalance in the awarding of scholarships. We also conducted a simple logistic regression analysis to examine the relationship between race and the probability that a student would qualify for a scholarship.

The third research question – whether the Michigan program is likely to meet the legislated goal of increasing access to higher education in the state – is answered by comparing the recipients of the first round of scholarship awards to patterns of college participation in the state.

## Results

### *Scholarship Qualification*

Just over 124,000 students in the high school graduating class of 2000 took the MEAP HST tests at some point before graduation. The distribution of these students is shown in Table 1.

Michigan is primarily a biracial state; almost 90 percent of the students taking at least one of the HST tests in 1999 were White or African American (for students who reported their race in the test file). While Whites represented 73 percent of all students taking at least one test (column B in Table 1), they made up almost 88 percent (column F) of all scholarship qualifiers. African American students were the most under-represented; while 17 percent of all test takers were African American, they qualified for only 4 percent of the scholarships. These results are not surprising given the evidence of the relationship between race and standardized test scores that we described earlier.

**TABLE 1**  
**Scholarship Qualification Rates**

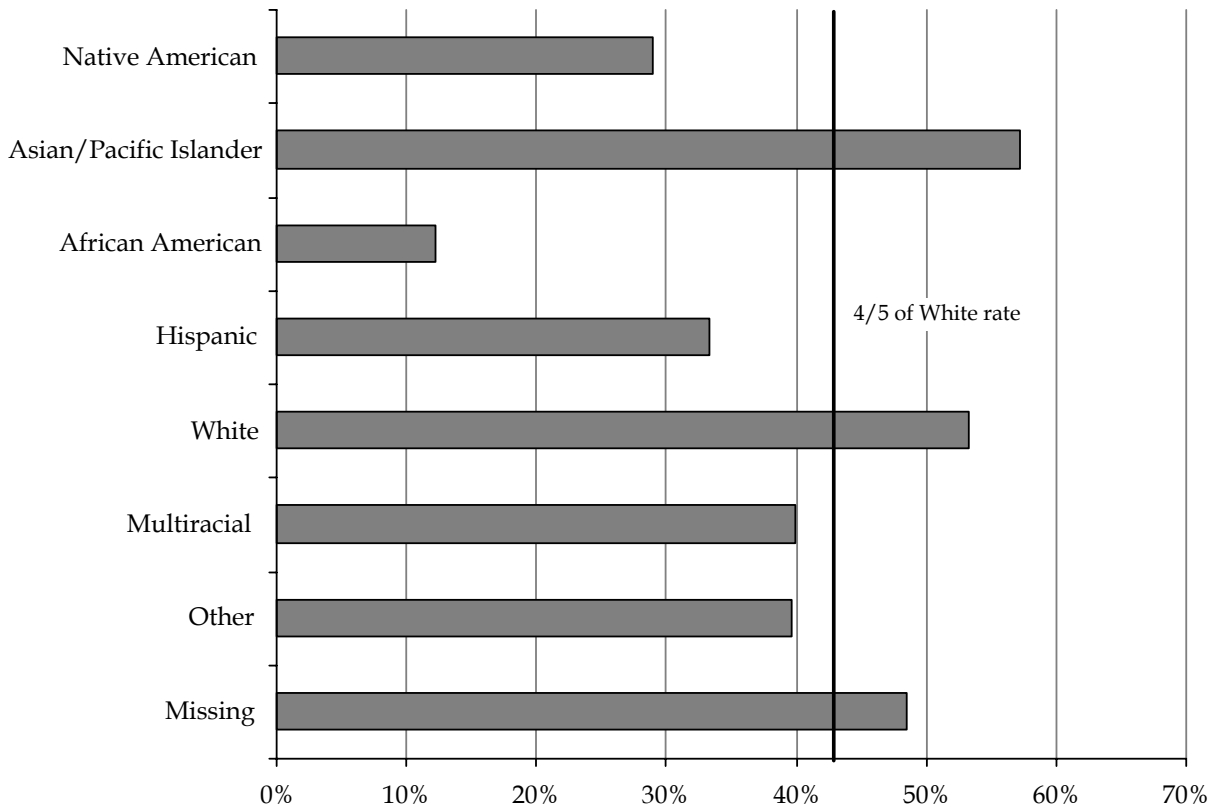
Group	A	B	C	D	E	F	G	H
	Total # of students	Group as % of all students*	# taking all 4 HST tests	% of all students taking all 4 HST tests (C/A)	# of students qualifying for a scholarship	Group qualifiers as % of all qualifiers*	% of total qualifying for scholarship (E/A)	% of those taking all 4 tests qualifying for scholarship (E/C)
Native American	970	1.1%	756	77.9%	219	0.6%	22.6%	29.0%
Asian/Pacific Islander	1,887	2.1%	1,684	89.2%	964	2.7%	51.1%	57.2%
African American	15,050	16.7%	9,961	66.2%	1,217	3.5%	8.1%	12.2%
Hispanic	2,361	2.6%	1,804	76.4%	601	1.7%	25.5%	33.3%
White	65,839	72.9%	57,620	87.5%	30,729	87.6%	46.7%	53.3%
Multiracial	1,885	2.1%	1,501	79.6%	599	1.7%	31.8%	39.9%
Other	2,379	2.6%	1,882	79.1%	745	2.1%	31.3%	39.6%
Missing	33,648	–	15,761	46.8%	7,635	–	22.7%	48.4%
Total	124,019	100.0%	90,969	73.4%	42,709	100.0%	34.4%	46.9%

\* Only those students with known race included in total.

Note: Column percentages may not sum to 100% due to rounding.

Another method of comparing the differences among racial groups is to examine the scholarship qualification rates. Keep in mind that a student had to have taken all four of the MEAP HST tests in order to qualify for a scholarship, even if she wished to pursue the alternative qualification path (via a combination of the MEAP tests and SAT, ACT, or WorkKeys test scores). White and Asian/Pacific Islander students were most likely to have taken all four tests, and thus started along the path necessary to qualify for a scholarship. Native American, African American and Hispanic students were least likely to have taken this initial step.<sup>5</sup>

Among those students who *did* take this first step toward qualifying for a scholarship, there was a large difference in the rate at which each racial group qualified for a scholarship. These differences can be seen in Figure 1. While over 50 percent of White and Asian American students qualified for scholarships (among those who took all four tests), less than 34 percent of Native American, African American and Hispanic students qualified. Also shown in Figure 1 is a line designating the “4/5ths” disparate impact test applied by the courts. Four-fifths of the White (the majority group) qualification rate is 42.7 percent; the qualification rates of all three minority groups fall well below this level.



**Figure 1**  
**Scholarship Qualification Rates by Race**

In order to apply the first test used by the courts in assessing disparate impact, the statistical difference test, we conducted a simple logistic regression analysis to look at the predicted probability that students of different races would qualify for a scholarship. Outcomes that are dichotomous in nature are generally analyzed using logistic regression.<sup>6</sup> We express the relationship between race and the probability of receiving a scholarship as a Delta-*p* statistic, recommended by Cabrera (1994) and Petersen (1985) as a method for expressing the relationship between a unit change in a predictor and the estimated percentage change in the outcome. For example, a Delta-*p*

value of 0.025 indicates that the predicted probability of receiving a scholarship for a student of that race is 2.5 percentage points greater than that of White students, the referent group. Table 2 presents the results of the logistic regression model. African Americans had a predicted probability of receiving a scholarship that was 37 percentage points less than that of Whites. All groups, with the exception of Asian/Pacific Islanders, had probabilities below that of Whites. The racial effect was statistically significant for all groups at a rate of  $p \leq .01$ .

**Table 2**  
**Predicted Probabilities of Qualifying for a Scholarship (Delta- $p$  statistic)**

Race	Delta- $p$
Native American	-0.229
Asian/Pacific Islander	0.039*
African American	-0.372
Hispanic	-0.190
Multiracial	-0.130
Other	-0.133
Race missing	-0.048

For all estimates,  $p \leq .001$  except \* $p \leq .01$

Table 3 presents the paths followed by students in different racial groups to qualify for the scholarships. Over 93 percent of students qualified via the four MEAP tests alone. Fewer than 2,800 students statewide, less than 7 percent of all receiving scholarships, qualified via a combination of MEAP and the two national college assessment tests, with no students qualifying via the WorkKeys test.

**Table 3**  
**Paths to Scholarship Qualification**

Group	# of students qualifying for a scholarship	<u>MEAP Tests Alone</u>		<u>MEAP and SAT</u>		<u>MEAP and ACT</u>		<u>MEAP and WorkKeys</u>	
		# qualifying	% qualifying	# qualifying	% qualifying	# qualifying	% qualifying	# qualifying	% qualifying
Native American	219	209	95.4%	1	0.5%	9	4.1%	0	0.0%
Asian/Pacific Islander	964	892	92.5%	10	1.0%	62	6.4%	0	0.0%
African American	1,217	1,143	93.9%	4	0.3%	70	5.8%	0	0.0%
Hispanic	601	568	94.5%	0	0.0%	33	5.5%	0	0.0%
White	30,729	28,689	93.4%	33	0.1%	2,007	6.5%	0	0.0%
Multiracial	599	557	93.0%	0	0.0%	42	7.0%	0	0.0%
Other	745	665	89.3%	4	0.5%	76	10.2%	0	0.0%
Missing	7,635	7,132	93.4%	10	0.1%	493	6.5%	0	0.0%
<b>Total</b>	<b>42,709</b>	<b>39,855</b>	<b>93.3%</b>	<b>62</b>	<b>0.2%</b>	<b>2,792</b>	<b>6.5%</b>	<b>0</b>	<b>0.0%</b>

*Assessing the Impact on College Participation Rates in Michigan*

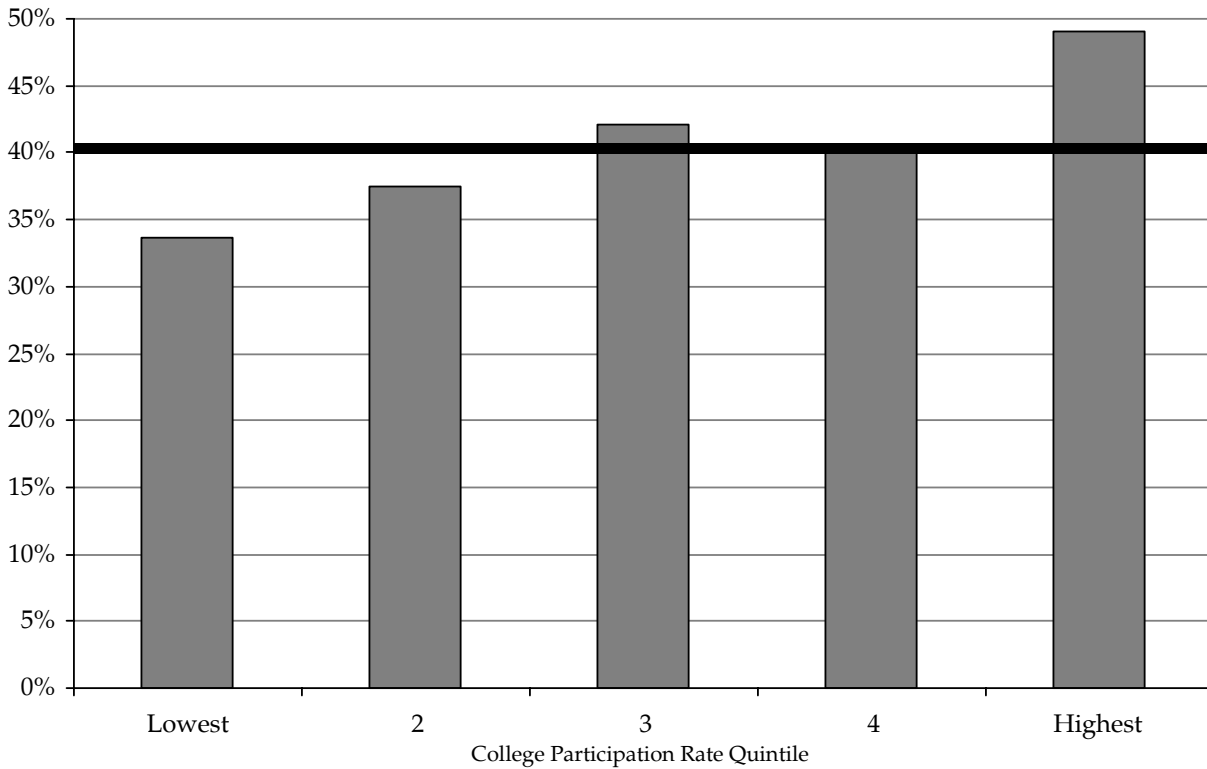
The Michigan Department of Treasury has not tracked the college enrollment status of the first cohort of students eligible for the Michigan Merit Award Scholarships. It does know, however, whether students who qualified for the scholarship attended college in the fall following their high school graduation, as students were required to provide the name of the school they were attending in order to receive the grant. In the fall of 2000, however, only 29,371 of the 42,709 scholarship qualifiers (69 percent) were enrolled in postsecondary education and had notified the Treasury of their intent to use the scholarship (Merit Award Program, 2001).<sup>7</sup>

Although the state has no such mechanism for tracking the college enrollment status of non-scholarship recipients, the Michigan Department of Education surveys public high schools in the state on the number of graduates each year and, among those graduates, how many enrolled in postsecondary education the subsequent fall (Michigan Department of Education, 1999). We used survey data from the 1997, 1998, and 1999 graduating classes to estimate the average postsecondary attendance rate in each high school.<sup>8</sup> We used data from all public high schools in the state that 1) reported data on the number of graduates attending postsecondary education, 2) had an average of at least 50 graduates during those three years, and 3) had at least fifty students who had attempted to qualify for the scholarships in the first eligible cohort. This provided a sample of 399 high schools in the state, representing approximately 46

percent of all public high schools, 75 percent of the graduating seniors, and 71 percent of the students attempting to qualify for the scholarships.

The correlation between the school's average college attendance rate over the three previous years (1997 – 1999) and the scholarship qualification rate of the school in 2000 is 0.332 ( $p \leq .001$ ), indicating that the schools that already sent higher percentages of students on to college had higher scholarship qualification rates.<sup>9</sup> To further explore this relationship, we divided the high schools into quintiles, based on their average college attendance rates over the previous three years. We then compared the scholarship qualification rates of each group. The results of this analysis are shown in Figure 2.

The fact that this analysis of the college-going rates of high schools focuses only on public school students has little relevance in assessing the likely impact of the scholarship program on postsecondary attendance in the state. Ninety percent of the scholarship qualifiers attended public schools; only 4,400 scholarship qualifiers attended private schools. In addition, private school students have been found to have higher college-going rates than students in public schools. In a study of students in the National Education Longitudinal Study of 1988, researchers found that students in private schools were more likely to attend college within a year of high school graduation, and were also more likely to be enrolled in college full-time (Sanderson, Dugoni, Rasinski, & Taylor, 1996).



The heavy horizontal line indicates the average scholarship qualification rate for all schools in the sample

**Figure 2**  
**Scholarship Qualification Rates by College Participation Rate Quintile**

### Discussion

#### *High-Stakes Testing and Disparate Impact*

The statistical results presented here speak directly to the first step of proving a disparate impact case of racial discrimination in educational testing under the implementing regulations of Title VI of the 1964 Civil Rights Act, as established by *Sharif v. New York State Education Department* (1989). The results demonstrate that there

are statistically significant differences in the rate at which students qualify for the Michigan Merit Award Scholarships, and that these differences are correlated with student race. In particular, we have shown that students from African American, Hispanic and Native American groups are disproportionately underrepresented among scholarship qualifiers relative to their representation among test-takers. The statistical findings meet both the first judicial test – a qualification rate for minority students that is statistically different from what would be expected from a random distribution – and the second judicial test – the minority group qualifying at a rate that is less than 80% of the majority group’s success rate. Moreover, both of these criteria are met even under the *Wards Cove* standard of comparing only the relevant pools of qualified applicants -- those students who took all four MEAP subject tests. These findings are consistent with the prior research on standardized tests and the demographic characteristics of the students taking them.

It is worth noting that the likely disparate impact of the Merit Scholarship program was known to the state at the time it was enacted. The state had for a number of years published statewide MEAP scores for students in different racial groups, data that led one report to conclude that “A white student is five times more likely than a black student to receive one of Michigan’s proposed \$2,500 college scholarships” (Egan, 1999, p. 6A). During the floor debate on the bill that created the program, the racial disparities in MEAP scores were discussed and debated (Michigan Senate, 1999, p. 957).

Although the main focus of our research was to establish the statistical evidence for the first step of proving a disparate impact case under Title VI, the remaining two

steps appear to be easily met by the plaintiffs in *White et al. v. Engler et al.* The second step requires the defendants to show that the use of the test, although disparately impacting minority groups, is nonetheless an educationally justified criterion, given the intended purposes of the program. Yet, the MEAP test was never validated as a measure of individual student achievement, the construct that the Merit Scholarships Act intended to reward. Indeed, the test was designed not to measure individual performance at all, but rather to rate school district performance in meeting curricular standards. Given this fact, the score cutoffs used to determine eligibility for the Merit Scholarships become arbitrary cutoffs, at best. Thus, there can be little legitimate educational justification for the use of the test to reward individual student achievement.

In addition, the Michigan Merit Award Program's use of the MEAP tests is educationally suspect because it conflicts with standards on the use of testing in education promulgated by a group of professional associations (American Educational Research Association, American Psychological Association, & National Council on Measurement in Education, 1999). For example, the AERA's Standard 1.4 states: "If a test is used in a way that has not been validated, it is incumbent on the user to justify the new use, collecting new evidence if necessary" (p. 18). The MEAP HST tests have never been validated as an instrument for use in awarding college scholarships, or for any type of postsecondary activity. The state's own guidelines indicate that "MEAP scores should not be used to rank students, schools, or districts" (Michigan State Board

of Education, 1998, p. 32). By using the MEAP scores to award scholarships to some students and not others, the state is in fact doing just that.

Similarly, Standard 13.7 states that: “In educational settings, a decision or characterization that will have major impact on a students should not be made on the basis of a single test score” (p. 146). While the scholarship program provides an alternative path for qualification by combining scores on the MEAP tests with other national standardized tests (SAT, ACT, and WorkKeys), the MEAP tests were the sole path to qualification for over 93 percent of all students (see Table 3). It is unknown why more students did not qualify via the alternative criteria, but it is evident that the MEAP tests are the *de facto* single criterion for qualification. This standard was echoed in a recent report issued by the Committee on Appropriate Test Use of the National Research Council (Heubert & Hauser, 1999), which recommended that “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 the student’s knowledge and skills, such as grades, teacher recommendations, and extenuating circumstances” (p. 279). While some may argue whether the awarding of a \$2,500 scholarship constitutes a “high-stakes decision,” it is hard to dispute that the receipt of such a scholarship can easily mean the difference between attending a postsecondary institution or not for a student from a poor family. Thus, there can be little legitimate educational justification for the use of the MEAP test to award scholarships for postsecondary attendance.

Finally, in the third step of a disparate impact case it is possible for plaintiffs to prevail over defendant's demonstration of educational necessity if plaintiff can show that an equally effective alternative method exists and has a less discriminatory impact. It is clear that reasonable other measures are available, on which the state of Michigan could base its awards of merit scholarships, and that would be less discriminatory than the current method of MEAP score cutoffs alone. These measures include high school GPA, class rank or standing, financial need, or any other combination of race-neutral "achievement" criteria. In both *Sharif* and *Cureton v. NCAA* (1999) the courts have recognized combinations of high school GPA and standardized test scores as less discriminatory alternatives to test scores alone for measuring academic achievement. These other measures, therefore, could easily satisfy the state's intent of rewarding achievement.

### *Race and College Participation*

We turn now to our last research question, whether the scholarship program is likely to achieve the legislated goal of increasing access to higher education in the state. The data on the racial distribution of the scholarship awards can be used to examine the likely impact of the program on college attendance in the state. The body of research that has examined the relationship between race and postsecondary educational attainment is very clear.<sup>10</sup> African American and Hispanic students enroll in and graduate from college at lower rates than do White and Asian American students.

These findings are generally consistent across different institutional sectors, over time, and in different regions of the country.

While there is limited evidence on educational participation and attainment rates in individual states, there is some to indicate that the patterns in Michigan mirror at least in part the broader national trends. Analyses of data from the Current Population Survey, conducted by the U.S. Bureau of the Census, indicate that there is a gap in the college completion rates of White, African American, and Hispanics in Michigan (Day & Curry, 1998). While 22.8 percent of Whites had completed a bachelor's degree or more in 1998, only 11.6 percent of African Americans and 7.7 percent of Hispanics had attained the same level of postsecondary education.<sup>11</sup>

In conjunction with a published study examining the college participation rates of students from different racial groups, Heller (1999) calculated the college participation rates of these groups in each state using enrollment data from the Integrated Postsecondary Education Data System, and population data from the U.S. Census Bureau. He found that in Michigan, the undergraduate college participation rates of Whites was 50 percent higher and Asian Americans 100 percent greater than that of both African Americans and Hispanics (personal communication). These differences in college participation rates mirrored the national trends.

Our results suggest, therefore, that the Michigan Merit Scholarships are not hitting the right target for the goal of increasing access to college in the state. By disproportionately awarding the scholarships to White and Asian American students, the program is giving funds to those who already enjoy the highest college attendance

rates, and thus where they are least likely to have a significant impact on access. At the same time, by disproportionately denying funds to African American, Hispanic and Native American students, those with the lowest existing college attendance rates, the program is missing the greatest opportunity for increasing overall access to college in the state.

We have shown that the Merit Program's use of MEAP tests has a negative disparate impact on minority students, cannot be justified educationally, and could easily be replaced by a more justifiable method that is less discriminatory. On top of all of that, it does not even appear to do a good job of addressing the legislatively mandated goal of increasing access to higher education in the state. Whether the court, in *White et al. v. Engler et al.*, will agree, however, is another matter altogether.

## References

- (2001). *ACT WorkKeys Assessments*, [<http://www.act.org/workkeys/assess/index.html>]. Iowa City, IA: ACT, Inc.
- American Educational Research Association, American Psychological Association, & National Council on Measurement in Education. (1999). *Standards for educational and psychological testing*. Washington, DC: American Educational Research Association.
- Bruschi, B. A., & Anderson, B. T. (1994). *Gender and ethnic differences in science achievement of nine-, thirteen-, and seventeen-year-old students*. Paper presented at the Annual Meeting of the Eastern Educational Research Association, Sarasota, FL (ERIC No. ED382751).
- Cabrera, A. F. (1994). Logistic regression analysis in higher education: An applied perspective. In J. C. Smart (Ed.), *Higher education: Handbook of theory and research* (Vol. X, pp. 225-256). New York: Agathon Press.
- Civil rights – Disparate-impact doctrine. (1990). *Harvard Law Review*, 103, 806-811.
- Cureton v. NCAA, 37 F. Supp.2d 687 (E.D. Pa. 1999).
- Day, J. C., & Curry, A. E. (1998). *Educational attainment in the United States: March 1998 (update)* (P20-513). Washington, DC: United States Bureau of the Census.
- Donahue, P. L., Voekl, K. E., Campbell, J. R., & Mazzeo, J. (1999). *NAEP 1998 reading report card for the nation and states* (NCES 1999-500). Washington, DC: U.S. Department of Education, National Center for Education Statistics.

- Egan, P. (1999, May 30). Critics concerned about racial fairness. *Lansing State Journal*, pp. 6A.
- Engler, J. (1999). *1999 State of the state address*. Lansing: State of Michigan, Governor's Office.
- Greenwald, E. A., Persky, H. R., Campbell, J. R., & Mazzeo, J. (1999). *NAEP 1998 writing: Report card for the nation and the states* (NCES 1999-462). Washington, DC: U.S. Department of Education, National Center for Education Statistics.
- Griggs v. Duke Power Co., 91 S. Ct. 849 (1971).
- Groves v. Alabama State Bd. of Educ., 776 F. Supp. 1518, 1531 (M.D. Ala. 1991).
- Hedges, L. V. (1999). Changes in the black-white gap in achievement test scores. *Sociology of Education*, 72(April), 111-135.
- Heller, D. E. (1999). Racial equity in college participation: African American students in the United States. *Review of African American Education*, 1(1), 5-29.
- Heller, D. E. (in press). The policy shift in state financial aid programs. In J. C. Smart (Ed.), *Higher education: Handbook of theory and research* (Vol. 17). New York: Agathon Press.
- Heubert, J. P., & Hauser, R. M. (Eds.). (1999). *High stakes: Testing for tracking, promotion, and graduation*. Washington, DC: National Academy Press.
- Kennedy, P. (1992). *A guide to econometrics*. Cambridge, MA: The MIT Press.
- Kimmel, A. P. (2001). Standardized tests: Low marks for fairness. *Trial*, 37(February), 41-45.

- Kleinbaum, D. G., Kupper, L. L., & Muller, K. E. (1988). *Applied regression analysis and other multivariate methods* (2nd ed.). Boston: PWS-KENT Publishing Company.
- Koretz, D. (1990). *Trends in the postsecondary enrollment of minorities*. Santa Monica, CA: The RAND Corporation.
- Lee, J. (1998). State policy correlates of the achievement gap among racial and social groups. *Studies in Educational Evaluation*, 24(2), 137-152.
- Merit Award Program. (2001). *Class of 2000 Michigan Merit Award recipients - Colleges and universities attended*. Lansing: Michigan Department of Treasury.
- Michigan Department of Education. (1999). *1999 Michigan school report*, [http://www.mde.state.mi.us/reports/msr/]. Lansing: Author.
- Michigan merit award scholarship act, 390.1451 (1999).
- Michigan Senate. (1999). *Journal of the Senate*, 54. Lansing: Author.
- Michigan State Board of Education. (1998). *Michigan Educational Assessment Program handbook, grades 4, 5, 7 and 8: 1997-1998 results*. Lansing: Author.
- Mumper, M. (1999, November). *HOPE and its critics: Sorting out the competing claims about Georgia's HOPE scholarship*. Paper presented at the annual meeting of the Association for the Study of Higher Education, San Antonio, TX.
- Orfield, G., & Wald, J. (2000, June 15). Testing, testing: The high-stakes testing mania hurts poor and minority students the most. *The Nation*, 270(22), pp. 38-40.
- O'Sullivan, C. Y., Reese, C. M., & Mazzeo, J. (1997). *NAEP 1996 science: Report card for the nation and the states* (NCES 97-497). Washington, DC: U.S. Department of Education, National Center for Education Statistics.

- Petersen, T. (1985). A comment on presenting results from logit and probit models. *American Sociological Review*, 50(1), 130-131.
- Reese, C. M., Miller, K. E., Mazzeo, J., & Dossey, J. A. (1997). *NAEP 1996 mathematics report card for the nation and states: Findings from the National Assessment of Educational Progress* (NCES 97-488). Washington, DC: U.S. Department of Education, National Center for Education Statistics.
- Sanderson, A., Dugoni, B., Rasinski, K., & Taylor, J. (1996). *National Education Longitudinal Study: 1988-1994: Descriptive summary report* (NCES 96-175). Washington, DC: U.S. Department of Education, National Center for Education Statistics.
- Sedlacek, D. A. (1995). *Using HLM and NAEP data to explore school correlates of 1990 mathematics and geometry achievement in grades 4, 8, 12 - Methodology and results* (NCES 95-697). Washington, DC: U.S. Department of Education, National Center for Education Statistics.
- Sharif v. New York State Education Department, 709 F. Supp. 345 (S.D.N.Y. 1989).
- Urdan, T., & Davis, H. (1998, February). *Differences by race and grade level in motivation for taking standardized achievement tests*. Paper presented at the bi-annual meeting of the Society for Research on Adolescence, San Diego, CA
- Vaseleck, J. (1994). Stop working and put down your pencils: The use and misuse of standardized admission tests. *Journal of College and University Law*, 20, 405-415.
- Wards Cove Packing Co. v. Antonio, 109 S. Ct. 2115 (1989).

## Notes

---

- <sup>1</sup> The proposal for the HOPE program was introduced during the first gubernatorial campaign in 1990 of former Governor Zell Miller of Georgia, and was widely seen as a key factor in his victory (Mumper, 1999).
- <sup>2</sup> For the first cohort of students, the qualifying SAT combined score is 1170 and the ACT composite score is 24. The WorkKeys test assesses individuals' knowledge of workplace skills, and is often taken by students enrolled in vocational programs in secondary school. It tests skills in the areas of applied mathematics, applied technology, listening, locating information, observation, reading for information, teamwork, and writing (ACT WorkKeys Assessments, 2001). Students who qualify for a scholarship via the WorkKeys test can use the funds for postsecondary vocational or technical training only.
- <sup>3</sup> The full \$2,500 scholarship is awarded for students attending college or some other form of postsecondary training in Michigan. Students attending out-of-state institutions are eligible for a \$1,000 award. In addition to these awards, students achieving certain scores on the 7<sup>th</sup> and 8<sup>th</sup> grade MEAP tests are eligible for up to an additional \$500 in scholarship funding.
- <sup>4</sup> The trial is expected to start later in 2001 (No. 00-CV-72882, E.D. Mich., filed June 27, 2000).
- <sup>5</sup> All group differences reported here were significant at a rate of  $p \leq 0.05$ .

- 6 For more information about the application of logistic regression, see Cabrera (1994), Kennedy (1992), and Kleinbaum, Kupper, and Muller (1988).
- 7 Students have up to seven years after high school graduation to claim the scholarship award, and it can be used for any level of postsecondary education.
- 8 These classes all graduated before enactment of the scholarship program, so its existence could not have influenced the college attendance rates in those years.
- 9 The analyses were weighted by the number of graduating seniors in each high school.
- 10 See for example Heller (1999) and Koretz (1990).
- 11 These figures are for all residents age 25 years old and over in the state. The Current Population Survey does not contain a large enough sample to calculate (with sufficiently small standard errors) educational attainment rates for racial/ethnic *and* age group subpopulations in each state. However, the national figures do show large gaps in the attainment (and college enrollment) rates for these groups and across different age cohorts. There is no evidence to indicate that Michigan's pattern is substantially different from the national trends.