



*CUNY's Testing Program: Characteristics,
Results, and Implications for Policy and
Research*

Stephen P. Klein and Maria Orlando

DRR-2047-1

5/99

*The Mayor's Advisory Task Force on the City University
of New York*

The RAND unrestricted draft series is intended to transmit preliminary results of RAND research. Unrestricted drafts have not been formally reviewed or edited. The views and conclusions expressed are tentative. A draft should not be cited or quoted without permission of the author, unless the preface grants such permission.

DRAFT

**CUNY's Testing Program:
Characteristics, Results, and Implications for Policy and Research.**

Stephen P. Klein, Ph.D. and Maria Orlando, Ph.D.

RAND

May 7, 1999

A Report Prepared for:

The Mayor's Task Force on the City University of New York

Preface

This research was conducted for the Mayor's Advisory Task Force on the City University of New York (CUNY), an advisory group established by New York City Mayor Rudolph W. Giuliani in May 1998. The Task Force is charged with reviewing, examining and making recommendations regarding: (1) the uses of City funding by CUNY, (2) the effects of open admissions and remedial education on CUNY, and on CUNY's capacity to provide college-level courses and curricula of high quality to its students, (3) the best means of arranging for third-parties to provide remediation services to ensure that prospective CUNY students can perform college-level work prior to their admission to CUNY, and (4) the implementation of other reform measures as may be appropriate.

This draft report examines the tests CUNY gives to its freshmen, as well as the demographic characteristics and general academic ability of CUNY's incoming freshmen. Other draft reports produced for the Task Force include:

- *The Governance of the City University of New York: A System at Odds with Itself*, Brian Gill, RAND DRR-2053-1
- *CUNY Statistical Profile 1980-1998, Volume 1: Draft Report*, Mary Kim, RAND DRR-2054-2, *Volume II: Databook*, Mary Kim, RAND DRR-2054/1-2
- *Financing Remediation at CUNY on a Performance Basis: A Proposal*, Arthur M. Hauptman, RAND DRR-2055-1

The RAND study was designed to provide the Mayor's Advisory Task Force the information and analysis they need to make recommendations to the Mayor on the future course of CUNY.

Overview

The colleges within the CUNY system have different admissions standards. In 1997, for example, York accepted entering freshmen into its bachelor's program who met one or more of the following four criteria: (1) Scholastic Assessment Test (SAT) total scores (i.e., math + verbal) ≥ 1020 on the "recentered" scale, (2) College Admissions Average (CAA) ≥ 75 with at least one math unit and 10 total units, (3) at least one math unit and 16 total units, or (4) a GED score of 300 or higher. Other CUNY colleges had higher and sometimes considerably higher standards for all or most of their students (e.g., such as if the school had affirmative action policies). In 1995, CUNY Board Regulations 15 and 16 stipulated that remediation at the senior colleges that do not offer Associate degree programs must be limited to one year or less beginning in the Fall of 1996.

CUNY uses three tests to decide which incoming students need to take remedial courses. These tests, which are collectively referred to as the Freshman Skills Assessment Tests or FSATs, are designed to measure basic reading, math, and writing skills.

This report has five sections. These sections evaluate the technical quality of the CUNY testing program, provide descriptive data on the academic ability of CUNY students (as indicated by their high school grades and College Board scores), describe the major features of a few research studies that could provide important information about the effectiveness of CUNY's remedial and regular educational programs, and discuss some policy and research options.

Specifically, Part I presents our findings regarding the reliability, validity, fairness, and costs of the FSATs. This analysis is especially critical of the Writing Assessment Test (WAT) portion of the FSATs because the WAT relies solely on a single essay question to make important decisions about individual students and this test has the lowest pass rate.

Part II describes the demographic characteristics and general academic ability of CUNY's freshmen as indicated by their actual (or imputed) SAT scores. These data show there are large differences among the CUNY colleges that grant bachelor degrees (and among those granting associate degrees) in the academic ability of their students. There also are large differences in average academic ability among racial/ethnic groups even after controlling on whether the student's primary language is or is not English. We then consider the likely consequences of these differences on the racial/ethnic composition of entering classes if CUNY raised its admission standards.

Part III shows that high school grades and test scores are fairly good predictors of grades at CUNY, especially for bachelor students. Part IV discusses some concerns about CUNY's database and additional studies that could be conducted. Part V discusses some policy options. An appendix contains supporting data from our statistical analyses.

PART I - ANALYSIS OF THE CUNY TESTING PROGRAM

The CUNY testing program uses three tests to determine which students require remedial instruction. The Reading Assessment Test (RAT) contains 45 multiple choice questions. The form currently being used by CUNY has a passing score of 30. The Mathematics Assessment Test (MAT) has two sections. The first section is used for making remedial placement decisions. It has 40 multiple choice items. A score of 25 or higher on this section is required for passing. There are seven forms of this section of the MAT (but each school decides which form to give when). On the Writing Assessment Test (WAT), students are given a choice between two questions to answer. They have 50 minutes to respond to the question they pick. Two new questions are asked each time the test is given. Two readers grade each answer on a 6-point scale. A total score of 8 or higher summed over the two readings is required for passing.

We obtained student demographic and academic information files from the CUNY Office of Institutional Research and Analysis. From these files, we identified 25,436 students who were first-time entering freshmen in Fall 1997. We deleted the 4% of these students (N=1,007) who did not have any FSAT scores on record. Thus, our analyses are based on the remaining 24,429 students who had at least one FSAT score on file.

Table 1 shows the mean score, standard deviation, and percent passing on each test in the cohort of incoming students in 1997. In this cohort, 65% of the students seeking a bachelors degree failed at least one test, 38% failed at least two, and 13% failed all three. The corresponding percentages among students seeking an associate degree were: 88%, 68%, and 37%. Only 35% of the bachelor students and 12% of the associate students passed all three tests.

The remainder of this section discusses the reliability, validity, fairness, and costs of the three CUNY tests. We also contrast these characteristics with those of the SATs.

Table 1
MEAN, STANDARD DEVIATION, AND PERCENT PASSING IN FALL 1997

Test	Bachelor (N = 8,705)			Associate (N = 15,493)		
	Mean	SD	% Pass	Mean	SD	% Pass
RAT	31.23	7.60	62	26.68	8.09	39
MAT	29.20	6.92	74	22.20	7.74	39
WAT	6.83	1.60	48	6.10	1.63	29
Total	67.32	12.97	35	55.45	13.51	12

Note: Mean total does not equal sum of RAT, MAT, and WAT scores due to missing data. Total passing equals the percentage of students passing all three tests.

Reliability

Reliability is usually reported on a scale from 0-1.00, where the higher the number, the greater the degree to which an individual student's relative standing (e.g., percentile rank) on one form of the test is consistent with that student's standing on another form of that test. However, in the context of how CUNY uses the FSATs, namely to make pass/fail decisions, reliability can be thought of as the likelihood that a student's pass/fail status on a test would remain the same regardless of which form of that test the student took. For example, would the student's pass/fail status on the WAT be affected if that student was asked an essay question that was administered to incoming freshmen in the fall of 1997 versus a question that was asked in the fall of 1998? Similarly, would a student's pass/fail status on the MAT depend on which form of that test the student took?

Determining the reliability of a test score typically involves measuring the consistency in student performance across the test's questions. In general, longer tests (as measured by the number of questions asked) have higher score reliabilities than shorter tests. Our computations of the reliability of the RAT and MAT were based on the scores earned by a sample of 1997 incoming freshmen (first time takers). These analyses found that both of these tests had a reliability of .89, which is reasonably high for tests of this length. For example, the estimated reliability for the somewhat longer SAT-M and SAT-V range from .91 - .94.¹

Several factors (besides the number of questions each student answers) influence the reliability of essay tests. One of the factors is the extent to which different graders assign the same score to an answer. This is called "inter-reader consistency." Another factor is the degree to which the questions measure the same thing. For example, one essay question might require a narrative response while another might require a persuasive argument. With respect to inter-reader consistency, studies done by CUNY suggest that about 15% of the students would have their pass/fail status affected on the WAT if a different reader graded their answers.²

It was not possible to determine the score reliability of the WAT because each student answers only one essay question. However, an estimate based on studies of similar single-question essay tests would be in the range of .25 to .60.³ The one essay question that is asked is in the persuasive genre. No other genres are tested.

Table 2 shows the relationship among the following three factors: (1) the reliability of a test's scores (on the 0-1.00 scale), (2) the passing rate of the test (i.e., from 0% to 100%), and (3) the misclassification rate. In this context, the misclassification rate is the percentage of students whose pass/fail status on one version of a test would be different from their pass/fail status on another version of that test (i.e., the percentage of students

¹ College Board (1999). *Counselors Handbook for the SAT program 1998-1999*. Author, New York, NY.

² Office of Academic Affairs (1998). *The CUNY Writing Assessment Test: Audit Results 1988-97*.

³ Dunbar, S., Koretz, D. and Hoover, H. D. (1991). Quality control in the development and use of performance assessment. *Applied Measurement in Education*, 4 (4), pp. 289-303.

who would be erroneously classified as passing or failing). The values in Table 2 were computed by Prof. David Freedman of the Statistics Department at the University of California at Berkeley.

Table 2
 PERCENTAGE OF STUDENTS WHOSE PASS/FAIL STATUS WOULD BE
 MISCLASSIFIED AT VARIOUS COMBINATIONS OF PASSING RATE
 AND SCORE RELIABILITY

Percent Passing	Score Reliability									
	.00	.10	.20	.30	.40	.50	.60	.70	.80	.90
90	19	17	17	16	14	13	12	11	9	6
80	32	30	28	27	25	23	20	17	14	10
70	42	39	37	35	31	29	25	22	17	12
60	48	45	42	39	36	32	29	25	20	14
50	50	47	44	40	37	33	30	26	21	14
40	48	45	42	39	36	32	29	25	20	14
30	42	40	38	35	32	29	26	22	18	13
20	32	30	28	27	25	23	20	18	14	10
10	19	18	16	16	15	13	12	11	09	6

Boxed area is probable range of pass/fail classification errors for the WAT.

About 60% to 70% of the bachelor students and 40% of the associate students pass the RAT and MAT (see Table 1). Both of these tests have reliabilities close to .90. Thus, each of these tests would misclassify only about 14% of the students. About 50% of the bachelor students and 30% of the associate students pass the WAT. The boxed area in Table 2 shows the probable range of misclassification rates for this test (i.e., assuming its reliability falls somewhere between .30 and .60). The misclassification rates in this zone are about 35% for bachelor students and 25% for associate students. In short, at least 25% (but probably more) of first time WAT takers are erroneously categorized; i.e., they fail when they should pass or pass when they should fail.

The low score reliability of a single question essay test stems mainly from students not being highly consistent with themselves in their writing ability across questions. In other words, a student's score is as much or more a function of the student's unique response to the particular question that is asked as it is of that student's overall ability to write. This means that a student might pass the essay question asked on the 1997 WAT but fail the one asked on the 1998 WAT while another student could easily have the opposite experience. The limited data that are available regarding inter-reader consistency on the WAT suggest that readers generally agree with each other in the score they assign to an answer. Hence, inter-reader consistency is probably not a major source of score reliability problems.

Validity

The validity of the FSATs for making initial placement decisions is measured by how well they distinguish between the students who truly do and do not need remedial instruction. To be valid, scores must first be reliable. Thus, reliability is a necessary but not sufficient condition for validity (which is why we are so concerned about the WAT). However, reliability alone does not insure validity. Scores must also reflect the abilities the tests are designed to measure.

We examined two indicators of the FSAT's validity, namely: (1) how well the scores on these tests correlate with scores on other similar and dissimilar tests (this is a type of "construct" validity) and (2) how well the FSAT scores predict a student's grade point average (GPA) at CUNY (this is called "predictive" validity). An appropriate measure of a student's success in remedial programs was not available.

Construct Validity. We obtained Scholastic Assessment Test (SAT) scores for 5,153 (59%) of the 8,705 entering bachelor's degree students and for 3,632 (23%) of the 15,493 entering associate's degree students in 1997. Overall, about 36% of the CUNY students had SAT scores. Roughly half of these scores were for students who had asked ETS to send their scores to CUNY. The other half were obtained from the College Board Corporation as part of a special study (CUNY does not require students to take the SATs and students may have to pay a nominal fee to have their SAT scores sent to a college).

The pattern of correlations among SAT and FSAT scores is consistent with what would be expected if the FSATs measured what they purported to measure. Table 3 shows that SAT-Verbal (SAT-V) scores correlated higher with RAT scores than with MAT scores while the reverse was true for SAT-Mathematics (SAT-M) scores. These findings provide some support for the construct validity of the RAT and MAT; i.e., these tests appear to measure reading and mathematical skills, respectively. It is not clear why the correlation between SAT-V and SAT-M scores was higher than the correlation between RAT and MAT scores, particularly since there was some restriction in the range of SAT scores (the more able students, as indicated by their FSATs, were somewhat more likely to take the SATs than other students).

Table 3
CORRELATION BETWEEN SAT AND FSAT SCORES

Correlation between	Bachelor Degree	Associate Degree	Total
SAT-V & RAT	.68	.53	.65
SAT-M & MAT	.66	.50	.65
SAT-V & MAT	.36	.23	.38
SAT-M & RAT	.41	.26	.42
SAT-V & SAT-M	.57	.50	.59
RAT & MAT	.38	.34	.42

Predictive Validity. We explored the predictive validity of the FSAT and SAT scores by assessing how well these scores correlated with the students' grade point average (GPA) at CUNY. Table 4 shows the mean validity coefficients (weighted by the number of students) across the nine senior colleges with at least 50 bachelor degree students and the ten colleges that had at least 50 associate degree students. The total FSAT score in these analyses is the sum of the student's RAT, MAT, and WAT scores. All of these validity coefficients have a possible range from -1 to 1, with -1 indicating a perfect negative relationship (i.e., as test score increases GPA decreases), 0 indicating no relationship, and 1 indicating a perfect positive relationship (i.e., as test score increases GPA increases).

Nationally, the typical correlation of SAT scores with freshman year GPA is about .50⁴ (out of a possible 1.00) after adjustment for restriction in range. However, the appropriateness of this adjustment has been questioned in the literature.⁵ If we had adjusted the values in Table 4 for bachelor students, the coefficients for SAT-V and SAT-M would be .22 and .38, respectively; i.e., well below the national average.

Table 4
MEAN CORRELATION WITH FRESHMEN GPA AT CUNY BY DEGREE TYPE

Degree Type	SAT Scores			FSAT Scores			
	Verbal	Math	Total	RAT	MAT	WAT	Total
Bachelor	.19	.24	.25	.18	.25	.14	.25
Associate	.08	.19	.17	.06	.19	.04	.16

⁴ College Board (1999). *Counselors Handbook for the SAT program 1998-1999*. Author, New York, NY.

⁵ Crocker and Algina (1986). *Introduction to Classical Modern Test Theory*. p. 226-227.

Table 5 shows the predictive validity coefficients separately by school and degree sought (bachelor or associate). We did this because of what appeared to be fairly large differences in grading standards among colleges. Specifically, there were large differences in average FSAT and SAT scores among colleges. These differences presumably reflect large differences in the average general academic abilities of their students. However, these differences did not correspond to the differences in these colleges' mean average first year grades (see Table 8 for details).

Table 5
CORRELATIONS WITH 1997-98 FRESHMEN GPA AT CUNY
BY DEGREE SOUGHT AND SCHOOL

Degree/College*	N	SAT Scores			FSAT Scores			
		Verbal	Math	Total	RAT	MAT	WAT	Total
<u>Bachelor</u>								
Baruch	945	.28	.22	.29	.34	.24	.21	.37
Brooklyn	1222	.27	.33	.33	.21	.32	.16	.30
City College	844	.18	.23	.24	.12	.21	.17	.20
Hunter	1550	.10	.23	.19	.08	.27	.06	.20
John Jay	807	.09	.14	.13	.09	.18	.09	.16
Lehman	639	.14	.21	.20	.17	.22	.12	.24
Queens	1096	.21	.23	.26	.21	.22	.16	.27
Staten Island	230	.24	.24	.30	.28	.32	.21	.39
York	414	.20	.29	.30	.14	.28	.10	.25
<u>Associate</u>								
Bronx	874	.14	.26	.22	.12	.26	.08	.21
Hostos	543	-.02	.09	.15	.00	.11	.08	.18
John Jay	594	.03	.15	.10	.04	.17	.05	.13
Kingsborough	1618	.19	.26	.28	.21	.29	.12	.31
La Guardia **	1884	-.03	.17	.07	-.03	.20	-.04	.09
Manhattan	2483	.01	.19	.11	.00	.19	.00	.11
Medgar Evers	423	.16	.23	.25	.18	.26	.14	.27
NYC Technical	1738	.03	.17	.11	.02	.15	.02	.10
Queensborough	1341	.11	.19	.18	.11	.20	.05	.19
Staten Island	1131	.22	.18	.25	.23	.22	.18	.29

* Results are reported for each degree/college combination with over 50 students.

** Correlations based on fall semester GPA only because spring data were missing.

The validity coefficients in Tables 4 and 5 are fairly low, especially for associate students. These coefficients may be depressed because of reliability problems with the students' GPAs. Specifically, many freshmen had their GPAs based on just a few courses. Officially, part timers comprised 9% of the bachelor students and 14% of the associate students.

Despite these modest part timer rates, half of the bachelor students had their first year GPAs based on less than 22 credit hours (about seven courses). Half of the freshmen associate students had less than 14 credit hours. This situation probably stemmed from many students taking remedial courses that did not count in the computation of credit hours towards GPA. The modest validity coefficients in Tables 4 and 5 cannot be explained by curtailments in the range of ability of the students tested (e.g., the standard deviations of their SAT-V and SAT-M scores were close to those in the population of all takers nationally).

Pass/Fail Scores. The issue of determining appropriate passing scores is at best problematic. We are not alone in this judgment. The CUNY administration has been advised by others to study the appropriateness of the passing scores,⁶ but to our knowledge, it has not systematically done so. Valid use of tests requires meaningful and defensible passing scores. For example, although RAT scores have a low positive correlation with first semester GPA, there is no evidence that students who score below the passing score on the RAT are not prepared for credit-bearing college courses, and students who score at or above it are prepared.

The appropriateness of the cut score becomes even more serious when we consider the WAT. This test has the lowest technical quality, but the most real impact. Students who fail it are generally regarded by CUNY faculty as truly lacking in writing ability. In addition, more students fail this test than fail the RAT or MAT (see Table 1). It is possible that the passing scores on the FSATs separate students into two distinct groups in terms of readiness for college courses. However, results of a pilot study conducted by the CUNY administration suggest that students with a score of 6 on the WAT may do as well in college courses as those who pass with a score of 8. Moreover, CUNY has not used any statistical methods to examine let alone control for the effect of varying difficulty in essay prompts from year to year. The same goes for possible differences over time in average reader leniency. Hence, it may be more difficult to earn an “8” one year than another.

The lack of controlled, systematic research into the appropriateness of the passing scores is a serious problem for the validity of the tests. Based on personal communications with CUNY faculty we have the impression that a number of small-scale cut score studies have been conducted throughout the history of CUNY’s testing program. Unfortunately, verbal descriptions and references to results of previous cut score studies cannot be taken at face value because little is known about the quality of the design and analyses in these studies. More research of this type must be performed and well-documented so that cut scores and their consequences can be examined.

⁶ Otheguy, R. (1990). The condition of Latinos in the City University of New York. A report to the Vice Chancellor for Academic Affairs and to the Puerto Rican Council on Higher Education. Unpublished report.

How Scores Are Used. The FSATs were originally intended to serve as a gate keeper to upper division courses at the senior level colleges. However, starting in about 1978, they were used system-wide as a mechanism for placing students in remedial courses. Some colleges also apparently use them to assess a student's progress in remediation, but this strategy may not be appropriate given the concerns that have been raised about the breaches in the security of the FSATs. CUNY was unable to provide validity evidence for any of the tests' purposes. Indeed, to our knowledge, the only systematic analysis of the validity of FSAT scores is contained in this report and our data apply only to their possible use as a predictor of first year grades. Our data do not speak directly to whether the FSAT scores are valid for the major purpose for which these scores are used, namely: deciding who needs remedial instruction.

Fairness

Several factors need to be considered in evaluating the fairness of a testing program. Our evaluation focused on racial/ethnic bias, test security, setting passing scores, and the decisions based on these scores. There are, of course, other fairness issues that we did not investigate. For example, we are not certain that the tests are appropriate for students who do not speak English as their primary language. We also did not explore whether the testing program was just as appropriate for older students as it was for younger ones. These and similar concerns can and should be addressed, but it was not possible for us to do so within the constraints of our study.

Racial/Ethnic Bias. We examined whether FSAT and SAT scores tended to over or under predict the grades of students in various racial/ethnic groups. In accordance with standard psychometric practice,⁷ we investigated this matter by using the data on all students to construct an equation to predict freshmen grades on the basis of test scores. For the reasons discussed in Part II of this report, this equation also included whether the student's primary language was or was not English. Next, we computed each student's "residual" score using the formula: Residual Score = Actual GPA - Predicted GPA.

In the context of this analysis, a test is considered "biased" against a group if its mean residual score is positive (i.e., if the mean of its actual observed GPA is greater than would be predicted on the basis of its placement test scores). A test would be "biased" in favor of a group if the opposite occurred, i.e., if its mean residual score was negative.

Our analyses found the FSATs and SATs were NOT biased against Hispanic or Black students. On both of these measures, the mean residual scores of the students in these groups were actually less than zero (see appendix for details). This finding indicates that on the average, the college grades of CUNY's Hispanic and Black students were lower than what would be expected on the basis of their FSAT and SAT scores. In contrast, Asian and white students tended to have positive residual scores. These results were obtained with both bachelor and associate degree students.

⁷ Standards for Educational and Psychological Testing (AERA, APA, NCME, 1985).

Test Security. We have been advised by reputable sources within and outside of CUNY (personal communications with College Board staff and CUNY administration, including P. Hasset, and L. Mirrer) that copies of the RAT and MAT can be purchased on the street. We had no way of determining the extent to which our population of first time takers had access to these tests. However, FSAT and SAT scores had similar validity coefficients. This finding suggests that security is not a serious problem for first time test takers. We suspect that a breach is likely to be more of a problem when higher stakes are attached to test outcomes, i.e. when the tests are used to decide whether a student has passed a remedial course or when they are used in the college admissions process.

Cost

We did not conduct an in-depth analysis of the costs of CUNY's testing program. We did find that CUNY spends about \$200,000 per year (\$8.35 per student) just to score the answers to the WAT (each answer is graded by two and sometimes three readers). This figure refers only to the initial administration of the WAT (i.e., it does not include grading of essays that might be administered after the completion of remedial coursework). This cost appears to be consistent with the cost of grading other essay exams (personal communication, Wayne Camara, The College Board) Eliminating the WAT may be opposed by the CUNY faculty who receive payment for grading the exam. Test development and administration are additional expenses for all the CUNY tests.

The SAT costs \$23.00 per student, which includes test development, administration and scoring. This fee also includes sending the results to four colleges. Fee waivers are available for students with financial need. Currently, about 35% of the incoming CUNY students already take the SATs (e.g., because they are applying to schools outside of CUNY).

Summary and Conclusions

We evaluated the technical quality of the CUNY testing program in the following four areas: reliability, validity, fairness, and cost. We found that score reliability was satisfactory on the RAT and MAT forms we analyzed. Assuming that the cut scores CUNY uses on these tests are appropriate, then each of them would misclassify the pass/fail status of about 14% of the students. About half of these misclassifications involve categorizing a student as a pass when that student should be failed while the other half are errors in the opposite direction.

We could not compute the reliability of the WAT because each student answers only one essay question. However, based on other research with essay tests, the WAT's reliability is probably in the .30 to .60 range. Consequently, the WAT misclassifies the pass/fail status of about 25% of the associate students and 35% of the bachelor students. These misclassification rates are a major concern because the WAT carries so much weight in

deciding who needs remedial instruction as a result of it having the lowest of the three FSAT passing rates.

FSAT scores (particularly the RAT and MAT) generally had low positive correlations with freshmen GPAs for bachelor students, indicating that these scores have reasonable predictive validity for these students. The predictive validity of the FSATs is comparable to that of the SATs in this sample. However, the FSATs and the SATs have very low predictive validities for associate students.

With respect to fairness, there does not appear to be any studies of the effectiveness of the FSATs for making placement decisions, which is the primary purpose of the tests, nor is there any empirical basis that we could find for the pass/fail cut scores. CUNY's practice of making important decisions based on a single score also is of some concern. There is no evidence of the tests being biased against African-American or Hispanic students.

Analysis of the test scores of 1997 entering freshmen did not suggest that the security of the tests had been breached, or at least not on a wide scale, because the correlation between FSATs and GPAs were comparable to the correlations between SATs and GPAs. We did not examine whether possible breaches in the RAT and MAT may have affected pass/fail decisions in remedial courses, i.e., when these tests are used in assessing whether a student has mastered a remedial course.

PART II – DEMOGRAPHICS, HIGH SCHOOL GRADES, AND SAT SCORES

The large proportion of CUNY students requiring remedial education has led to concerns about the overall academic ability of the CUNY students. This section provides descriptive information about CUNY students with regard to their demographic characteristics and academic ability (as measured by their SAT and FSAT scores, high school grades, and first year GPAs). We also provide information about the relationship between demographic groups and test scores throughout the CUNY colleges to explore the possible impact of changing policies regarding admission and remediation.

Demographics. In terms of racial/ethnic background, the four largest groups of students among fall 1997 entering freshmen at CUNY were as follows: Asians—10%, African-Americans—27% (herein-after referred to as Blacks), non-Hispanic whites—20% (herein after referred to as Whites), and Hispanics—28%. Almost all of the students in the “other” group were missing a valid racial/ethnic code (see Tables 6a and 6b).

About half of the entering freshmen (both bachelor and associate) said English was their primary language; i.e., they said they were native English speakers and/or preferred to speak in English. For the purposes of the analyses below, we classified these students as “English Speakers” and everyone else as “English Learners.” Some of the students in the latter category may in fact be fluent in English, but we had no way of identifying who

they were, and for them, English was a second language. Tables 6a and 6b also show that Hispanics had the largest number and percentage of English language learners.

Table 6a
 NUMBER OF STUDENTS IN EACH RACIAL/ETHNIC GROUP BY DEGREE
 SOUGHT AND WHETHER THEY ARE ENGLISH SPEAKERS OR LEARNERS

Group	Associate Degree			Bachelor Degree			Grand Total
	English Speakers	English Learners	Total	English Speakers	English Learners	Total	
Asian	221	961	1182	410	963	1373	2574
Black	2999	1720	4719	1129	629	1758	6520
Hispanic	1703	2848	4551	949	1242	2191	6777
White	1752	978	2730	1270	951	2221	4980
Other	1012	1299	2311	550	612	1162	3578
Total	7687	7806	15493	4308	4397	8705	24429

Table 6b
 PERCENTAGE OF STUDENTS IN EACH RACIAL/ETHNIC GROUP BY DEGREE
 SOUGHT AND WHETHER THEY ARE ENGLISH SPEAKERS OR LEARNERS

Group	Associate Degree			Bachelor Degree			Grand Total
	English Speakers	English Learners	Total	English Speakers	English Learners	Total	
Asian	3	12	8	10	22	16	10
Black	39	22	30	26	14	20	27
Hispanic	22	36	29	22	28	25	28
White	23	13	18	29	22	26	20
Other	13	17	15	13	14	13	15
Total	100	100	100	100	100	100	100

High School Grade Point Average (HSGPA). New York City public high schools graduated 29,203 students in June 1997. Of this group, 8,559 (29%) entered CUNY in the fall of 1997. The mean HSGPA of those who did and did not go to CUNY were 75.5 and 72.2, respectively (as computed by CUNY on a 0 to 100 point scale). The corresponding means among those who took the Regents exam were 76.3 and 77.3. These data indicate that the HSGPAs of the June 1997 high school graduates who went to CUNY were similar to the HSGPAs of the June 1997 graduates who were likely to be college bound but did not go to CUNY (the standard deviation was 15 points in the group taking the Regents exam that did not go to CUNY). CUNY is not drawing from just the bottom of the New York City pool of graduates. The mean HSGPAs of the June 1997 graduates enrolling in associate and bachelor programs at CUNY were 70.0 and 80.8, respectively. The corresponding means for the June 1997 graduates going to CUNY who took the Regents exams are 71.2 and 80.8.

Imputing SAT Scores. We conducted a separate analysis to estimate what the SAT scores at CUNY would be if all entering students took Part I of the SAT. This was done by calibrating the RAT and MAT scores to SAT-V and SAT-M scores, respectively; and the FSAT total (RAT+MAT+WAT) to the SAT total for the roughly 9,000 entering students in 1997 who had FSAT and SAT scores.⁸ For example, 5% of the 9,000 students had a RAT score of 14 or less and 5% had a SAT-V score of 260 or less. We therefore said a RAT score of 14 was “equivalent” to a SAT-V of 260. Similarly, we set a RAT score of 17 equivalent to an SAT-V of 310 because 10% of the students had a RAT score of 17 or less and 10% had an SAT-V of 310 or less (see Appendix for details).

We repeated the process above for every 5th percentile point to create an “equi-percentile cross-walk” between the two tests. We then used this cross-walk to construct a linear regression equation for imputing a student’s SAT score from that student’s corresponding FSAT score for each student who did not already have an SAT score. Because the FSAT total included the WAT score, the sum of a student’s imputed SAT-V and SAT-M scores did not always equal that student’s imputed SAT-Total score.

Finally, we ran two checks on the accuracy of the links, namely: (1) that there was a strong linear relationship between an imputed SAT score and its corresponding FSAT score and (2) that the cross-walk and regression equation was stable. We tested the stability of the equations by randomly splitting the sample in half according to month of birth (students born on odd-numbered months in one group, even-numbered months in the other group), and repeated the equating process separately on the two halves. Results were very similar to those obtained for the full sample. Visual inspection of the degree of linear relationship, combined with this stability check lead us to conclude that all three links (i.e., RAT to SAT-V, MAT to SAT-M, and FSAT total to SAT total) clearly passed both checks (see Appendix). Thus, we have a high degree of confidence in the accuracy of the links for the limited purpose of conducting the analyses described below.

⁸ The SAT-Total was equated to the total FSATs rather than only the RAT and MAT because the total FSAT scores correlated slightly higher with SAT total than did the combined RAT and MAT score.

Mean SAT Scores At CUNY. Our analyses used SAT scores on the recently “re-centered” scale. This scale has a national mean of 500 on each test (1,000 on total score). Table 7 shows the estimated mean SAT-V, SAT-M, and SAT-Total scores for all entering CUNY freshmen in 1997 by which degree they were seeking. One benchmark for interpreting SAT scores is the NCAA’s eligibility requirements for athletic scholarships, namely: a student must have an SAT total score of at least 820 (and a high school GPA of at least 2.5 in 13 core academic subjects).

Another benchmark is SAT scores at other colleges. Given CUNY’s relatively modest admissions standards, it is not surprising that its SAT scores are fairly low in comparison to most other colleges in New York and nationally. For example, we identified four New York colleges from the annual USNews college rankings that were in the same tier (Northern Universities, Tier 2) as Brooklyn, Baruch, and Hunter. The 25th and 75th percentiles of SAT Total scores for these four colleges, College of New Rochelle, Iona College, SUNY Plattsburgh, and SUNY Oswego, were 890-1050, 910-1100, 960-1140, and 980-1180 respectively. In comparison, the 25th and 75th percentile points for CUNY’s bachelors students were 795 and 1040.

Table 7
ESTIMATED MEAN SAT SCORES OF 1997 FRESHMEN

Test	Bachelor Degree	Associate Degree
Verbal	447	402
Math	469	402
Total	916	799

The mean total SAT scores of the June 1997 graduates who went to CUNY that fall were 817 for the 4,173 associate students and 910 for the 4,386 bachelor students. The corresponding means for other associate and bachelor students in this class were 796 and 926. These data indicate that the SAT scores of the 8,559 June 1997 graduates from New York City public schools who went to CUNY that fall were fairly comparable to the SAT scores of the other CUNY freshmen in this entering class.

Table 8 shows the mean SAT and FSAT scores at each college. Some schools (such as John Jay) are listed twice because they have large numbers of both bachelor and associate degree seeking students. Within a degree, schools are listed in descending order of their SAT total scores. This sequence is almost identical to the order of their mean total FSAT scores. Probably because of the reliability problems discussed above, the mean RAT and MAT scores tracked SAT scores across schools much better than did mean WAT scores.

The mean total FSAT scores in Table 8 may not equal the sum of the mean RAT, MAT and WAT scores because these means are based on slightly different numbers of students (i.e. not all students have scores for all three tests). Missing data on one or more tests also affected the mean SAT total score, which was estimated (using their total FSAT score) for students who did not take the SATs. As a result, the mean SAT total score may not equal the sum of the means of the SAT-V and the SAT-M. In most cases, the difference is negligible. At Hostos, however, only 50% of the students took all three parts of the FSAT, and these students tended to have higher RAT and MAT scores than other Hostos students. The estimated mean SAT total score for Hostos is based on the subset of students with complete data and is therefore not representative of the entire freshman class. A more appropriate estimate of the mean SAT total score for Hostos is therefore the sum of its mean SAT-V and SAT-M scores (i.e., 668 rather than 747).

There are large differences in the general academic ability (as measured by SATs and FSATs) of the students attending different schools within the CUNY system. The top six colleges in Table 8 have much more able students (as measured by FSAT and SAT total scores) than do the next three schools. For example, there is a very large (54-point) difference in mean SAT total scores between City College (the 6th school on the list) and John Jay (the 7th school). Similarly, Bronx, Hostos, and La Guardia had much lower mean SAT-Total scores than did other colleges, including other community colleges. Moreover, Staten Island's mean SAT score was substantially higher than the mean at the other community colleges granting associate degrees.

The large differences in mean student ability among schools (as measured by FSATs and SATs) do not correspond to differences in their grading standards. To illustrate, Table 8 shows that the mean GPA at Baruch, the school with the most academically able students, was lower than the mean GPA at six of the other senior colleges. It also was lower than the mean GPA at Hostos and at some of the other community colleges. These differences raise serious concerns about transferring grades and credits across CUNY's colleges.

Table 8
 MEAN SAT AND FSAT SCORES AND FRESMEN GPA BY DEGREE SOUGHT
 AND COLLEGE FOR STUDENTS ENTERING IN FALL 1997

Degree/College*	N	Mean GPA	SAT Scores			FSAT Scores			
			Verbal	Math	Total	RAT	MAT	WAT	Total
<u>Bachelor</u>									
Baruch	1082	2.33	464	501	968	33	32	7	72
Hunter	1712	2.50	460	483	946	33	31	7	71
Queens	1205	2.59	461	481	942	32	30	7	69
Staten Island	252	2.73	453	472	926	33	30	7	69
Brooklyn	1368	2.30	449	479	924	31	29	7	67
City College	954	2.46	440	479	918	30	30	6	67
John Jay	904	2.31	438	426	864	31	26	7	63
York	464	2.35	410	441	847	28	27	6	61
Lehman	711	2.37	407	410	811	28	25	6	59
<u>Associate</u>									
Staten Island	1440	2.22	438	422	859	31	24	7	61
Medgar Evers	527	2.11	415	403	810	28	22	6	56
Kingsborough	2030	2.40	411	404	809	28	22	6	56
Queensborough	1717	2.04	409	409	809	27	23	6	57
Manhattan	3044	2.37	405	411	808	27	23	6	56
NYC Technical	2170	2.13	400	408	800	27	23	6	56
John Jay	715	1.88	411	389	794	28	21	6	55
La Guardia	2076	2.45	390	398	776	25	22	6	53
Hostos	671	2.44	328	340	747	20	16	6	50
Bronx	1103	2.14	374	366	717	23	18	5	47

* Results are reported for each degree/college combination with over 50 students.

Relationship Between Student Demographics and SAT Scores. Figure 1 shows how SAT total scores of students seeking a bachelors degree are related to their racial/ethnic group and primary language (English Speakers versus Learners). Figure 2 shows the corresponding data for students seeking an associate degree. In both figures, each horizontal bar represents the middle 50% of the distribution of scores for a group. The left-hand side of each bar shows the 25th percentile within that group, the vertical line in the middle of the bar shows the 50th percentile point, and the right hand side of the bar shows the 75th percentile point. For example, the bottom bar in Figure 1 shows that roughly the middle 50% of the Black bachelor English Learners had an SAT total score between 740 and 950. The median (50th percentile point) in this group was just below 850. Figure 1 also shows that the middle 50% of the Asian bachelor English Speakers had SAT scores between 900 and 1140. The mean SAT score among all those who take the SAT nationally (i.e., among students who are aspiring to go to college) is about 1000. The mean among all bachelor students at CUNY was 916 (which is far below the national average).

Figures 1 and 2 show that within a racial/ethnic group, “English Speakers” generally had much higher scores than “English Learners” (as per the definitions of these groups used earlier in this report). When English fluency is held constant, Asian and White students generally had higher SAT total scores than their classmates (the bachelor and associate students in the “other” category had SAT score distributions that were comparable to the entire populations of bachelor and associate students, respectively; see the top bar in each figure). In fact, White and Asian English Learners generally had scores that were as high or higher than those of Hispanics and Blacks who were English Speakers.

A comparison of Figures 1 and 2 shows that bachelor students tended to earn substantially higher SAT scores than associate students. For example, the 25th percentile among all bachelor students corresponded to an SAT score of 795 which is exactly equal to the median (50th percentile) score for associate students. Moreover, the median score for bachelor students (920) corresponded to the 75th percentile for associate students.

The differences in the distributions of SAT scores between certain racial/ethnic groups are comparable in size to the differences between bachelor and associate students. For example, Figure 1 shows that among English Speakers, about 75% of the White and Asian bachelor students had higher SAT scores than half of the Black and Hispanic bachelor students. As discussed earlier in this report, this disparity is not due to the tests being biased against Blacks or Hispanics.

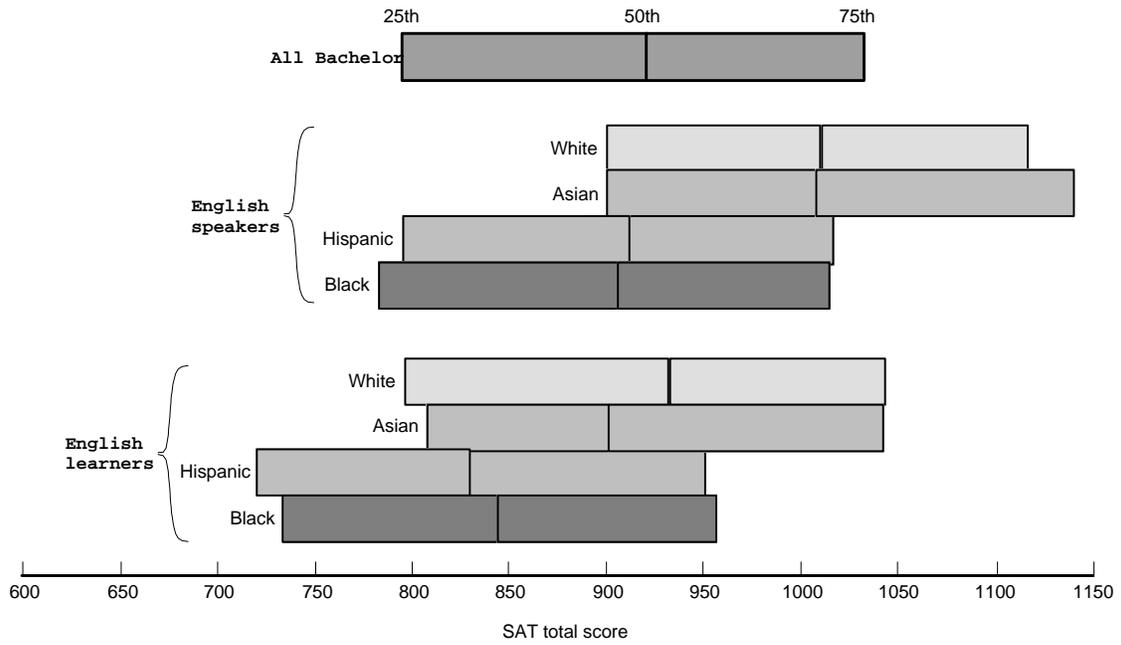


Figure 1. Interquartile Range by Racial/ethnic Group for Bachelor Students

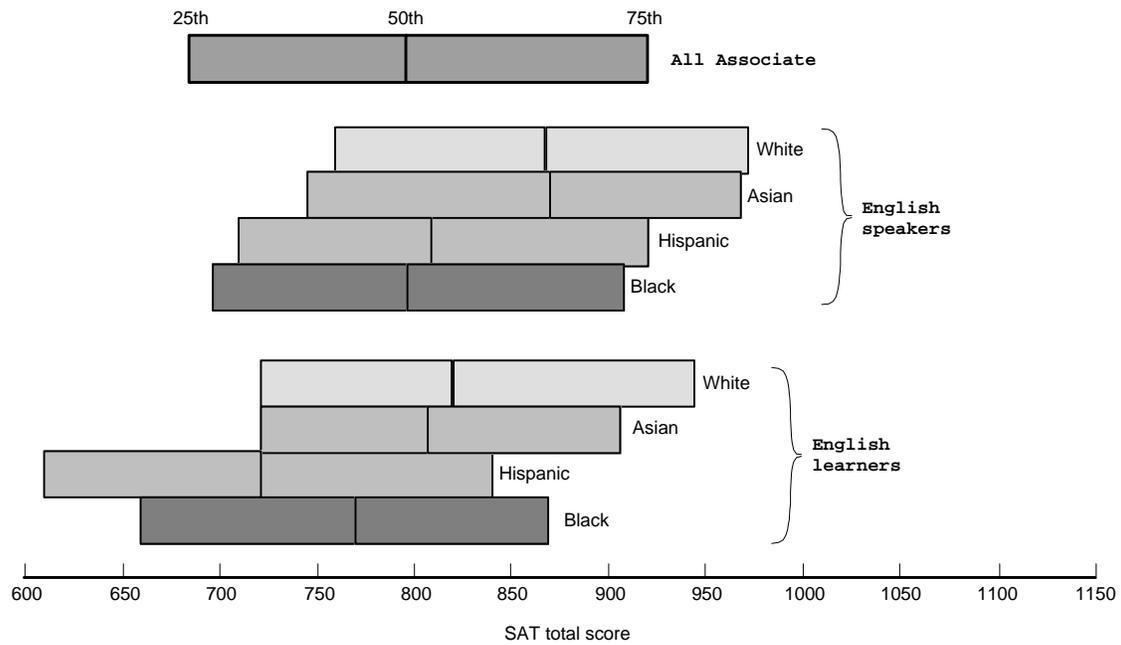


Figure 2. Interquartile Range by Racial/ethnic Group for Associate Students

Policy Implications of Differences Among Groups. Figure 1 shows that if CUNY raised its admission standards for bachelor students at some or all of its senior colleges, then the percentages of Black and Hispanic students who would reach these standards would most likely be lower than the percentages of Asian and White students who would meet them. Thus, at least in the *short term*, the data suggest that raising standards would result in the most selective schools having disproportionately fewer Black and Hispanic students than Asian and White students. This situation could, of course, be mitigated if CUNY adopted an affirmative action policy that involved imposing substantially higher standards for Whites and Asians than it employed for Blacks and Hispanics. We do not discuss in this report the public policy and political consequences of using different admission standards for different groups to insure racial balance in access to CUNY colleges.

It is difficult to predict the *long term* consequences of higher admission standards on racial disparities. For example, higher standards could lead to improved academic preparation of Black and Hispanic students (i.e., before they come to CUNY) which in turn could raise their college graduation rates. Thus, higher standards could lead to raising the net number of Blacks and Hispanics who graduate from CUNY. Shifting the policy focus from access to college to graduation rates could therefore lead to different decisions regarding the appropriateness of imposing higher admission standards on all students.

Finally, there is no other factor (besides racial/ethnic group) that can be inserted into the admissions process that will lead to racial/ethnic balance. For example, the admissions office at UCLA found that including a student's socioeconomic status in the admissions process will not come close to restoring the racial/ethnic balance that was achieved by the affirmative action policies that were in place prior to the implementation of Prop. 209; i.e., the proposition that eliminated racial/ethnic group from the admissions process (personal communication with W. Doby, Vice Chancellor). To achieve such balance, at least in the short run, the admissions process will have to consider the student's racial/ethnic group or radically change its admissions standards. There is no way around this. In addition, policy makers will have to develop guidelines for defining what constitutes "balance." For example, must a racial/ethnic group's share of the student body at a college equal its share among all high school graduates in New York City, among all CUNY students, etc.? If so, then this would essentially raise admission standards for Whites and Asians, but not for Blacks and Hispanics.

PART III – ANALYSIS OF HIGH SCHOOL DATA

Although there are over a million school children in New York City's public schools, only about 30,000 graduate from high school each spring. About 8,600 of these students (29%) went to CUNY and of this group (96%) took at least one Regents English or mathematics exam. The 18,551 non-CUNY bound June 1997 high school graduates who also took at least one Regents exam had a moderately (and statistically significantly) higher mean score on these exams than did those who went to CUNY. The difference was about one quarter of a standard deviation unit on each test. The gender and racial/ethnic composition of the CUNY bound students was very similar to the composition of the non-CUNY bound students. Taken together, these data suggest that the more able college bound high school graduates from New York City's public schools were somewhat more likely to go some place other than CUNY, but the difference was not dramatic.

Within CUNY itself, the students in the cohort of 8,559 spring 1997 high school graduates were much more likely to seek a bachelors degree than were the 15,870 fall 1997 CUNY freshmen who were not recent NYC public high school graduates. The percentages in these two groups were 51% and 21%, respectively. However, within a degree track, the spring 1997 and non-spring 1997 graduates had very similar test scores and demographic characteristics. For example, among those seeking a bachelor degree, their respective mean RAT scores were 31.1 and 31.4; their mean SAT-M scores were 466 and 472; and their corresponding percentages of "English Speakers" were 48% and 52%. This similarity suggests that once there is control on degree type, the relationship of CUNY GPAs to high school grades and test scores in the cohort of spring 1997 high school graduates is likely to be similar to the relationship between these variables in the population of all entering CUNY students (but this should be checked by further research because results could be influenced by factors that we were not able to control).

We used the cohort of 8,559 June 1997 New York Public High School graduates who went to CUNY that fall to examine how first year grades at CUNY were related to high school grades and to the scores on the English and mathematics portion of the New York State Regents exams (herein after referred to as "Regents"). This was done by constructing 12 regression equations for bachelor students and another 12 for associate students.

All 12 equations contained the same set of background and demographic characteristics, namely: racial/ethnic group, language (English Learner versus Speaker), and college. The latter variable was included to help compensate for differences in grading standards across schools within CUNY. The 12 models differed with respect to whether they included one or more of the following variables: SATs (i.e., SAT-V and SAT-M), FSATs (RAT, MAT, and WAT), high school grade point average (HSGPA), and score on the Regents English and math exam.

Table 9 shows the squared multiple correlation (R-square) for each model for each group (comparisons can be made between these data and those in Tables 4 and 5 by squaring the correlation coefficients in those tables). The R-square value is an index of the extent to which differences in first year grades among students can be explained by differences in their background characteristics and test scores (i.e., by the variables in the model). Specifically, an R-square value indicates the proportion of the variance in the students' grades that can be accounted for by the variance in these students' predictor scores.

Table 9
R-SQUARES OF VARIOUS MODELS IN PREDICTING FRESHMAN GPAs

Model Number	Variables in the Model	Bachelor (N = 4,429)	Associate (N = 4,069)
1	Covariates (School, Language, & Race)	.06	.06
2	Covariates + Regents	.14	.11
3	Covariates + HSGPA	.17	.07
4	Covariates + HSGPA + Regents	.20	.11
5	Covariates + SATs	.10	.07
6	Covariates + SATs + Regents	.14	.11
7	Covariates + SATs + HSGPA	.20	.08
8	Covariates + SATs + HSGPA + Regents	.21	.11
9	Covariates + FSATs	.12	.08
10	Covariates + FSATs + Regents	.15	.11
11	Covariates + FSATs + HSGPA	.21	.08
12	Covariates + FSATs + HSGPA + Regents	.22	.11

In general, the R-square for predicting freshmen grades from high school grades and admissions test scores is in the .10 to .15 range (prior to any adjustment for restriction in range). Values over .20 are definitely above average. Table 9 shows that combining Regents scores and/or HSGPAs with a student's FSAT or SAT scores yields a more accurate prediction of a student's likelihood of success at CUNY than does using any of these measures by themselves. In fact, the highest R-squares are obtained by combining HSGPA with two of the three sets of test scores. However, as we saw in Tables 4 and 5, the predictor variables are much more accurate in estimating first year grades for bachelor students than they are for predicting the grades of associate students.

PART IV - ADDITIONAL RESEARCH ACTIVITIES

This section notes some concerns we have with CUNY's database. We then discuss several research studies that would provide useful information if CUNY retains its current testing program and/or launches a new one.

Improve Data Quality. While conducting the analyses for Parts I and II of this report, we encountered some questionable data. For example, a large percentage of students were missing key demographic information and several students had more than 35 credits in their freshmen year (and one had 46 credits even though it is highly unlikely that a student took a dozen or more courses over two semesters).

In addition, our discussions with CUNY staff indicated that in the fall of 1997, two WAT forms were used (forms 33 and 34). However, the computer file for the freshmen entering in 1997 had codes for over 50 different forms! The number of students answering forms 33 and 34 were 7,422 and 5,320, respectively, out of 23,300 takers (1,129 students did not have a WAT score). These data suggest that only slightly more than 50% of the students took one of the two forms that were presumably administered to everyone. We do not know whether these results stem from clerical/key-entry errors, problems with the documentation for the electronic files we received, or whether they signal more significant and pervasive problems. Whatever the reason, it is evident CUNY needs to improve the quality of its student information system.

Analyze High School Data. The combination of SAT or FSAT scores with high school grades or Regents scores provides a more accurate prediction of a student's college grades than do any of these variables by themselves. In addition, measurement specialists generally recommend using more than one test score to make important decisions about individual students. In light of such considerations, we suggest that CUNY determine whether using high school data (including scores on statewide tests) would improve the assessment of a student's readiness for college level work at CUNY. We began to explore this matter in Part III of this report, but a more thorough analysis is required, particularly since we were only able to look at about one third of the CUNY freshmen.

Document Basis for Pass/Fail Standards. As noted in Part I of this report, there does not appear to be any documented empirical or theoretical basis for the passing scores CUNY selected for the RAT, MAT, and WAT. In addition, these tests have very different passing rates. Consequently, if these tests are to be retained, we strongly recommend that research be conducted to determine what the passing score on each test should be.

Assess the Consistency of CUNY's Grading Standards. The relatively low correlations of both SAT and FSAT scores with CUNY GPAs of Associate students may stem at least in part from problems with the grading system; i.e., the outcome variable may not be very reliable. If so, that could depress the correlation of CUNY GPAs with other measures. A study of the reliability of grades at CUNY could therefore isolate the source(s) of the low correlations between GPAs and test scores.

In addition, the data in Table 8 suggested that there were large differences in grading standards across CUNY's colleges (e.g., the schools with the most academically able students—as measured by FSAT and SAT scores—did not have the highest average GPAs). Research on CUNY's grading policies and practices would help to identify the sources of these inconsistencies and provide insights into how they could be eliminated so as to increase the fairness of the grades and facilitate their transfer across schools.

A more in-depth analysis also could examine the degree to which grading standards at CUNY are comparable to those at other colleges. This type of research might involve administering a common set of test questions as part of final course exams to students from different schools within and outside of CUNY, giving standardized high school advanced placement tests to CUNY students at the end of comparable first year courses, and similar strategies.

Evaluate the Effectiveness of Different Remedial Programs. Two out of three entering bachelor students and almost every freshman associate student receives remedial instruction at CUNY. Some of the remedial programs these students receive are no doubt more cost effective than others as indicated by the amount of time and other resources they require to help students reach the level of verbal and mathematical proficiency CUNY students need to do college work. Thus, it would be useful to determine which programs or program types are most effective for which types of students. This research will need to include some non-FSAT measures of student proficiency because many of the remedial courses now use the FSATs as part of their instructional program.

Develop Valid and Appropriate System-wide Measures of Student Abilities. There are certain basic reading, math, and writing skills that all CUNY graduates should master. That is why CUNY instituted the FSATs and according to its own standards on these tests, a very large percentage of its incoming students require remedial instruction. However, CUNY has no systematic way of assessing whether the remedial instruction that was given to these students was effective; i.e., whether its graduates actually possess the requisite skills.

CUNY's proposed "60th credit" (single prompt essay) exam will not assess mastery of the relevant basic abilities because it does not assess math or science skills. It also suffers from the same score reliability problems as the WAT. Hence, we suggest that CUNY consider developing or adopting a valid system of secure tests for assessing whether its students have acquired the basic skills that are commensurate with a bachelor and associate degree. We also suggest that CUNY begin formally monitoring and reporting upon the success of its graduates on relevant licensing and certification tests, such as for teachers and accountants, as one factor in assessing the quality of its instructional programs in these areas.

Examine the “Value Added” of a CUNY Degree. The “value added” of an institution of higher learning is measured by the degree to which its students are eventually substantially better off (in terms of income, job and life satisfaction, etc.) than are similarly situated individuals who did not go to CUNY. For example, does going to CUNY lead to securing a better job, becoming more productive, etc.? CUNY could answer these and related questions by conducting a longitudinal study of a stratified random sample of the students who enrolled in a given year (e.g., fall 1992) to find out what they are doing now, their thoughts about the quality of the education they received, and similar matters.

PART V - POLICY OPTIONS AND RECOMENDATIONS

CUNY must decide whether to maintain its generally modest admissions standards. If it retains these standards, then it will have to do things: (1) provide effective remedial instruction to large numbers of students at both the senior and community colleges and (2) have a defensible method for determining which students receive that instruction.

Another strategy would be to raise admission standards at some or all of the other senior colleges, and channel those students requiring remedial instruction to the community colleges and/or other public or private programs. The rationale for this strategy is that it would more efficiently serve the needs of students who need remedial assistance as well as raise academic standards. This will also increase the prestige of the CUNY system and thereby potentially attract more able students to its colleges. If this approach is adopted, then CUNY will need a valid and appropriate set of criteria for setting cut scores and determining which students should go to which schools.

The results presented in this report indicate that in deciding between these and other options, CUNY will need to keep in mind several factors, including the following:

It may not be appropriate for CUNY to continue to use the FSATs to make high stakes decisions, such as whether a student is required to take a remedial course or be admitted to a particular college. The major reasons for this concern are (1) the security of the RAT and MAT have been breached and (2) the score reliability of the WAT is far below what is appropriate for making important decisions about individual students. It is just not adequate for the task it is being asked to perform, especially since it is the major determiner of whether a student is required to take a remedial course.

Writing skills are certainly important to measure, but the WAT cannot be trusted to provide an accurate index of those skills. In addition, scoring costs alone on this test are about \$200,000 per year. Hence, if CUNY continues its FSAT program, then it should either (1) base the WAT score on several essay questions per student or (2) combine the WAT and RAT scores into a composite total language arts score. In addition, CUNY should go through a formal standard setting process and analysis to determine the appropriate passing (“cut”) score on each component test in the FSAT program.

If CUNY decides to impose stricter admission standards for bachelor students at some of its senior colleges, then at least in the short term, white and Asian students will have a much higher likelihood of being admitted than will students from most other racial/ethnic groups. These differences stem from Black and Hispanic students tending to have lower and sometimes substantially lower admissions credentials than their classmates (see Figures 1 and 2). These disparities are not due to problems in the tests. Specifically, our analyses found that the differences in average test scores between groups did not stem from gross differences in English fluency rates between groups or the tests being biased against Black or Hispanic students. In fact, we found that the tests actually favored these students in the sense that their actual GPAs at CUNY were statistically significantly lower than what would be predicted on the basis of their FSAT or SAT scores while the reverse was true for Asian and non-Hispanic white students.

Finally, Part IV of this report listed several areas in which CUNY might conduct additional research. These areas include conducting further investigations of the utility of using high school grades in making selection and placement decisions, assessing the reliability and appropriateness of CUNY's grading and curriculum standards, evaluating the effectiveness of various remedial programs for different types of students, instituting a quality control check on basic skills, and examining the value added of a CUNY degree

APPENDIX

Table A1
RESULTS OF RAT RELIABILITY ANALYSIS

 Cronbach Coefficient Alpha for RAW variables : 0.889312
 Cronbach Coefficient Alpha for STANDARDIZED variables: 0.891348

Deleted Variable	Raw Variables		Std. Variables	
	Correlation with Total	Alpha	Correlation with Total	Alpha
RNEW01	0.450136	0.885866	0.453901	0.887947
RNEW02	0.330599	0.887827	0.338122	0.889613
RNEW03	0.322419	0.887743	0.327721	0.889761
RNEW04	0.488465	0.885422	0.493883	0.887367
RNEW05	0.401629	0.886691	0.406865	0.888626
RNEW06	0.383578	0.886827	0.383626	0.888961
RNEW07	0.437624	0.886077	0.443585	0.888096
RNEW08	0.384987	0.886914	0.390559	0.888861
RNEW09	0.259312	0.888746	0.259190	0.890737
RNEW10	0.391690	0.886832	0.396778	0.888771
RNEW11	0.314483	0.887856	0.317349	0.889910
RNEW12	0.384706	0.886832	0.387304	0.888908
RNEW13	0.302946	0.887934	0.307172	0.890055
RNEW14	0.397982	0.886604	0.399352	0.888734
RNEW15	0.361581	0.887201	0.359543	0.889306
RNEW16	0.306394	0.888071	0.303542	0.890106
RNEW17	0.348162	0.887371	0.347825	0.889474
RNEW18	0.151740	0.890449	0.150164	0.892274
RNEW19	0.299077	0.888155	0.294670	0.890233
RNEW20	0.197278	0.889366	0.193508	0.891665
RNEW21	0.479729	0.885386	0.483231	0.887522
RNEW22	0.327247	0.887731	0.329356	0.889738
RNEW23	0.477501	0.885504	0.481734	0.887544
RNEW24	0.373307	0.887003	0.377173	0.889053
RNEW25	0.312349	0.887973	0.310439	0.890008
RNEW26	0.330806	0.887644	0.334679	0.889662
RNEW27	0.404070	0.886622	0.408807	0.888598
RNEW28	0.421097	0.886618	0.426745	0.888340
RNEW29	0.539278	0.884497	0.541899	0.886667
RNEW30	0.356283	0.887239	0.359808	0.889302
RNEW31	0.320769	0.887858	0.317870	0.889902
RNEW32	0.394750	0.886655	0.389253	0.888880
RNEW33	0.421566	0.886215	0.416735	0.888484
RNEW34	0.286281	0.888426	0.283855	0.890387
RNEW35	0.485592	0.885327	0.485980	0.887482
RNEW36	0.421284	0.886220	0.416976	0.888481
RNEW37	0.419806	0.886260	0.418705	0.888456
RNEW38	0.472804	0.885531	0.472755	0.887674
RNEW39	0.452649	0.885721	0.449718	0.888008
RNEW40	0.286339	0.888408	0.284085	0.890383
RNEW41	0.503616	0.884857	0.499181	0.887290
RNEW42	0.509447	0.884958	0.507955	0.887162
RNEW43	0.257573	0.888831	0.252443	0.890832
RNEW44	0.261633	0.888614	0.256485	0.890775
RNEW45	0.225542	0.889200	0.220645	0.891282

Table A2
RESULTS OF MAT RELIABILITY ANALYSIS

 Cronbach Coefficient Alpha for RAW variables : 0.894655
 Cronbach Coefficient Alpha for STANDARDIZED variables: 0.892351

Deleted Variable	Raw Variables		Std. Variables	
	Correlation with Total	Alpha	Correlation with Total	Alpha
MNEW01	0.235997	0.894527	0.238097	0.892253
MNEW02	0.063013	0.895595	0.065885	0.894863
MNEW03	0.340882	0.893050	0.344046	0.890621
MNEW04	0.515323	0.890316	0.512129	0.887993
MNEW05	0.372716	0.892608	0.377389	0.890104
MNEW06	0.382619	0.892475	0.382947	0.890017
MNEW07	0.251150	0.894625	0.255339	0.891989
MNEW08	0.472990	0.890994	0.468169	0.888685
MNEW09	0.559159	0.889778	0.558557	0.887258
MNEW10	0.422850	0.891812	0.423526	0.889384
MNEW11	0.376689	0.892554	0.372757	0.890176
MNEW12	0.290049	0.893792	0.292050	0.891424
MNEW13	0.278043	0.893873	0.284294	0.891544
MNEW14	0.330802	0.893145	0.334827	0.890764
MNEW15	0.362045	0.892743	0.362092	0.890341
MNEW16	0.284403	0.893770	0.289325	0.891466
MNEW17	0.370357	0.892680	0.369517	0.890226
MNEW18	0.193780	0.894535	0.199142	0.892848
MNEW19	0.511086	0.890346	0.506553	0.888081
MNEW20	0.347050	0.892947	0.353859	0.890469
MNEW21	0.565286	0.889549	0.562226	0.887199
MNEW22	0.447553	0.891436	0.447608	0.889007
MNEW23	0.548854	0.889933	0.547609	0.887431
MNEW24	0.401064	0.892172	0.401374	0.889730
MNEW25	0.547737	0.889891	0.544802	0.887476
MNEW26	0.464176	0.891163	0.462266	0.888778
MNEW27	0.427623	0.891734	0.425148	0.889359
MNEW28	0.569727	0.889554	0.565205	0.887152
MNEW29	0.283716	0.894072	0.281212	0.891591
MNEW30	0.468588	0.891056	0.462243	0.888778
MNEW31	0.396411	0.892557	0.398836	0.889770
MNEW32	0.516099	0.890288	0.511696	0.887999
MNEW33	0.389378	0.892340	0.391722	0.889881
MNEW34	0.427349	0.891739	0.422547	0.889400
MNEW35	0.317813	0.893533	0.317226	0.891036
MNEW36	0.437013	0.891579	0.432544	0.889243
MNEW37	0.484328	0.890801	0.480621	0.888489
MNEW38	0.437823	0.891565	0.432376	0.889246
MNEW39	0.229594	0.894942	0.226459	0.892431
MNEW40	0.362361	0.892796	0.357514	0.890412

Table A3
 MEAN FRESHMAN GPAs AND RESIDUAL SCORES
 BY RACIAL/ETHNIC GROUP AND DEGREE FOR FOUR MODELS

Degree/Ethnicity	N	Actual GPA	Mean Residual Score for:			
			FSAT Model 1	FSAT Model 2	FSAT Model 3	FSAT Model 4
<u>Bachelor</u>						
White	1940	2.61	.12**	.13**	.12**	.12**
Black	1477	2.31	-.04*	-.03	-.04	-.02
Hispanic	1834	2.23	-.16**	-.14**	-.15**	-.13**
Asian	1232	2.53	.06*	.01	.05*	.00
Other	900	2.48	.04	.04	.04	.05
<u>Associate</u>						
White	2206	2.43	.14**	.15**	.15**	.16**
Black	3821	2.13	-.08**	-.07**	-.08**	-.08**
Hispanic	3505	2.17	-.07**	-.06**	-.07**	-.05**
Asian	1018	2.56	.22**	.13**	.22**	.14**
Other	1835	2.28	.00	.00	.00	.00

* = significant at .05, ** = significant at .001

Notes: Residual score = Actual GPA – Predicted GPA. The R-Squares for models 1-4 were .07, .08, .07, and .08, respectively. The dependent variable for all four models was the students freshman GPA. In addition to degree, all four models also included dummy variables for primary language (English Speaker versus English Learner), and college. The models differed in terms of their other predictor variables as follows: Model 1 used FSAT total score (i.e. RAT + MAT + WAT); Model 2 used each of these three tests as separate variables; Model 3 used SAT total score; and Model 4 used SAT-V and SAT-M scores as separate variables. Where necessary, SAT scores were imputed from FSAT scores using the procedures described in Part II of this report. The tabled N's are for the FSAT models (and they are about 99% of the N's in the SAT models).

Table A4
FSAT – SAT CROSSWALK

percentile	FSATs	SAT-Tot	RAT	SAT-V	MAT	SAT-M
5	35	590	14	260	11	280
10	41	640	17	310	14	320
15	44	680	19	330	15	340
20	47	720	21	350	17	350
25	50	740	23	360	18	370
30	52	770	24	380	20	380
35	54	790	25	390	21	400
40	56	810	27	400	22	410
45	58	830	28	410	24	420
50	60	860	29	430	25	430
55	62	880	30	440	26	450
60	64	910	31	450	27	460
65	66	930	32	460	29	470
70	68	960	34	480	30	490
75	70	980	35	490	31	500
80	73	1020	36	510	33	520
85	75	1050	37	530	34	550
90	78	1100	39	560	35	580
95	82	1190	40	600	37	620

Figure A1
RELATIONSHIP BETWEEN RAT TOTAL SCORE AND SAT-V

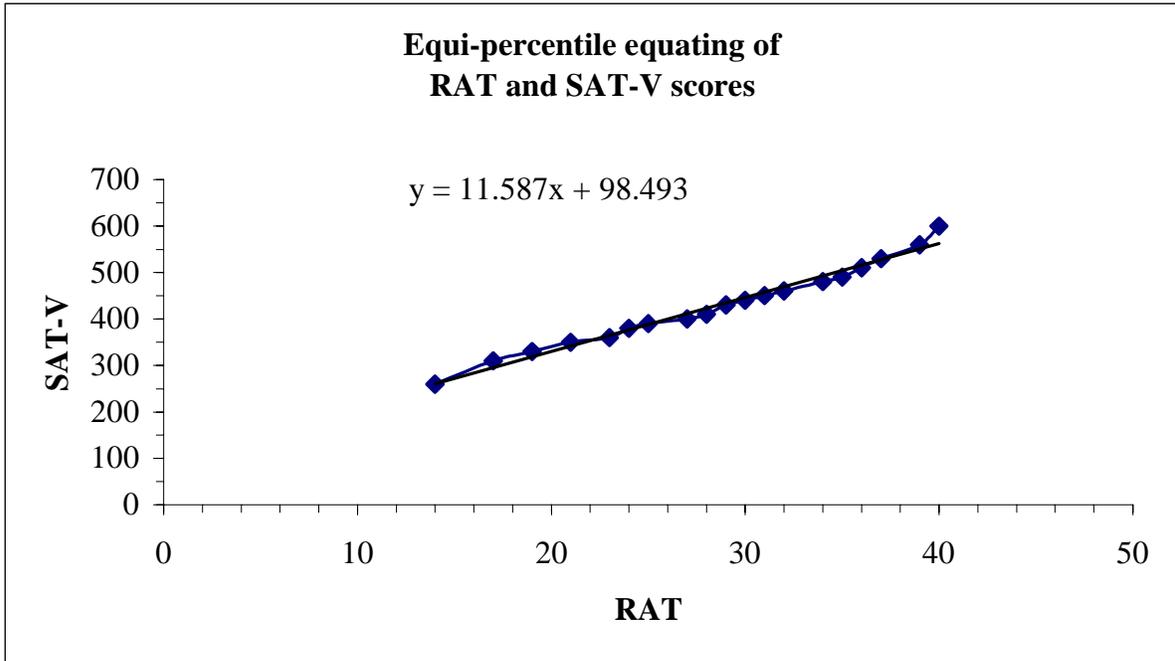


Figure A2
RELATIONSHIP BETWEEN MAT TOTAL SCORE AND SAT-M

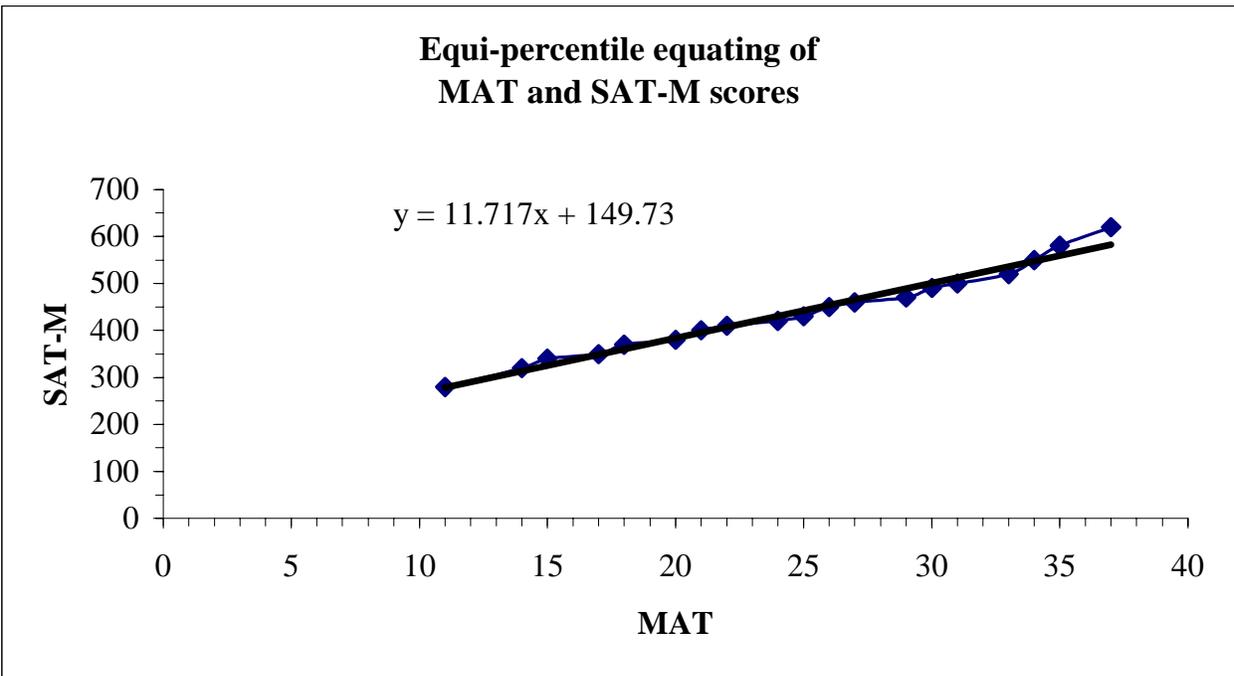


Figure A3
RELATIONSHIP BETWEEN FSAT TOTAL SCORE AND SAT TOTAL SCORE

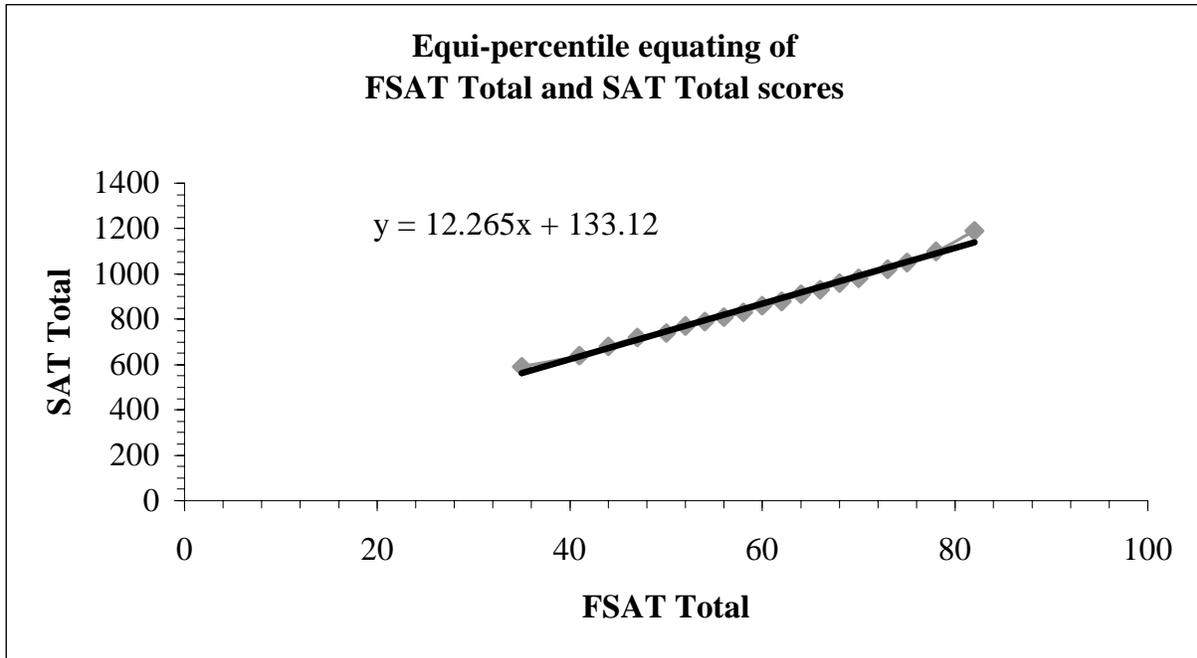


Table A5
CROSSWALK EQUATIONS FOR FULL SAMPLE AND RANDOM HALVES

	Crosswalk relationship		
	Verbal	Math	Total
Full Sample	$y = 11.6x + 98$	$y = 11.7x + 149$	$y = 12.3x + 133$
Half 1	$y = 11.6x + 98$	$y = 11.6x + 150$	$y = 12.0x + 145$
Half 2	$y = 11.8x + 94$	$y = 11.9x + 142$	$y = 12.2x + 137$